

# Level 3 Fake News Identification Project

---

## Big Data Summative Assignment

---

Author: Z0954757

## Environment Set-Up

---

This project was created in an environment with the following packages:

- Python 2.7,
- Numpy,
- Pandas,
- Matplotlib
- NLTK - and data installed by entering the python interpreter typing the following and installing the data associated with the NLTK book:

```
>> import nltk
```

```
>> nltk.download()
```

- spaCy
- Scikit learn
- Keras

## Shallow Learning approach

---

The shallow learning approach is encapsulated in `shallow.py`.

## Deep Learning approach

---

C) Feature Extraction:

The required word2vec function is available as `word2vec.py` and implements spaCy's prebuilt vectorizer.

D) i) The LSTM is encapsulated in `deep.py`

D) ii) Due to hardware and time limitations, there is no implementation for D) ii).

# Results presentation

---

The module `main.py` is included. Running `$ python main.py` will start a command line interface to select what code you wish to run. The layout of the function follows the order of the report.

It first asks if you would like to run the shallow approaches. Typing `y` will execute this step, or any other input not including `y` will skip it.

Next it asks if you would like to run the further investigations into optimising the the shallow learning models. If `y` is entered, you are then prompted if you would like to regenerate the data or load it from a saved csv file of the outcomes.

In both cases, it will then generate the contour plots for the grid search optimisation results.

## Report

---

The report is included as `report.pdf`.