Description : Develop a machine learning program to identify when an article might be fake news. Run by the UTK Machine Learning Club.

EVALUATION : The evaluation metric for this competition is accuracy, a very straightforward metric.

$$accuracy = correct\ predictions\ /correct\ predictions + incorrect\ predictions$$

Accuracy measures false positives and false negeatives equally, and really should only be used in simple cases and when classes are of (generally) equal class size

Submission Format

For every article in the test dataset, submission files should contain two columns: `id` and `label`. The `id` column should refer to a row in the `test.csv` file, and the `label` column should refer it's class of reliable (`0`), or potentially fake (`1`).

The file should contain a header and have the following format:

`id,label`

`182041,1`

`182042,0`

`182043,1`

`182044,0`

`etc.`

Data Description

train.csv: A full training dataset with the following attributes:

- id: unique id for a news article
- title: the title of a news article
- author: author of the news article

- text: the text of the article; could be incomplete
- label: a label that marks the article as potentially unreliable
    - 1: unreliable
    - 0: reliable

test.csv: A testing training dataset with all the same attributes at train.csv without the label.

submit.csv: A sample submission that you can