Assignment 2: Propaganda Detection

Due Date: March 30

In this assignment we will aim to develop a tool to detect propaganda from the input documents, the task is described in: https://propaganda.math.unipd.it/semeval2021task6/

1. Data set:

<https://github.com/di-dimitrov/SEMEVAL-2021-task6-corpus>

This link describes the detailed format and data split, please follow the split for training and testing your system.

2. (10pts) Baseline: Use BERT or any other basic embedding features to represent text to train a supervised model.

3. (5pts) Improve the baseline by methods such as: (1) generate training data automatically for data augmentation; or (2) incorporate more advanced linguistic features (e.g., semantic role labeling) or external knowledge (e.g., entity linking to background knowledge base); or (3) incorporate visual features.

4. (up to 5 extra points): Write detailed analysis and discussions on what worked and what did not, and propose potential solutions for the remaining challenges.

5. Submission: Please name your submission as `netid\_assignment2.zip`, including a report named as `netid\_report.pdf` and the code `netid\_code`. The code should include a README.md with environment and running instructions.

The submission is done through Blackboard Compass2g (https://compass2g.illinois.edu/webapps/blackboard/content/listContentEditable.jsp?content\_id=\_5671021\_1&course\_id=\_62362\_1 )