Case Study

March 2020 / 2021

Context

Fake news has been there since before the advent of the Internet. They are widely accepted to be fictitious articles deliberately fabricated to deceive readers. Social media and news outlets publish fake news to increase readership or as part of psychological warfare.

Objective

The aim of this Case Study is to construct a fake news detection algorithm. It should be able to detect forms of news consisting of deliberate disinformation via online social media.

Datasets

Each student will receive 02 datasets, a training set and a testing set. Datasets are not the same from one student to another one. Each dataset contains 05 variables:

- *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

Evaluation criteria

Your submission report should contain clear and well explained codes and their outputs. You should present clearly and briefly your:

- Methodology
- Data description
- Modelization steps
- Interpretation of results
- Recommendations / Insights

NB: Students are free to choose the software of their convenience.