

Fake News - Identify fake web profiles

Supervisor: Diego Esteves

Xiaotian Zhou
2743095



Idea

- Use the skill learned from Exercise 3 to:
- Build a fake news classifier which helps
- Identify contents from different web URLs
- Tools: Python Jupyter notebook
- Libraries: Scikits.learn, Pandas, [BeautifulSoup](#) (extract data from webpage)

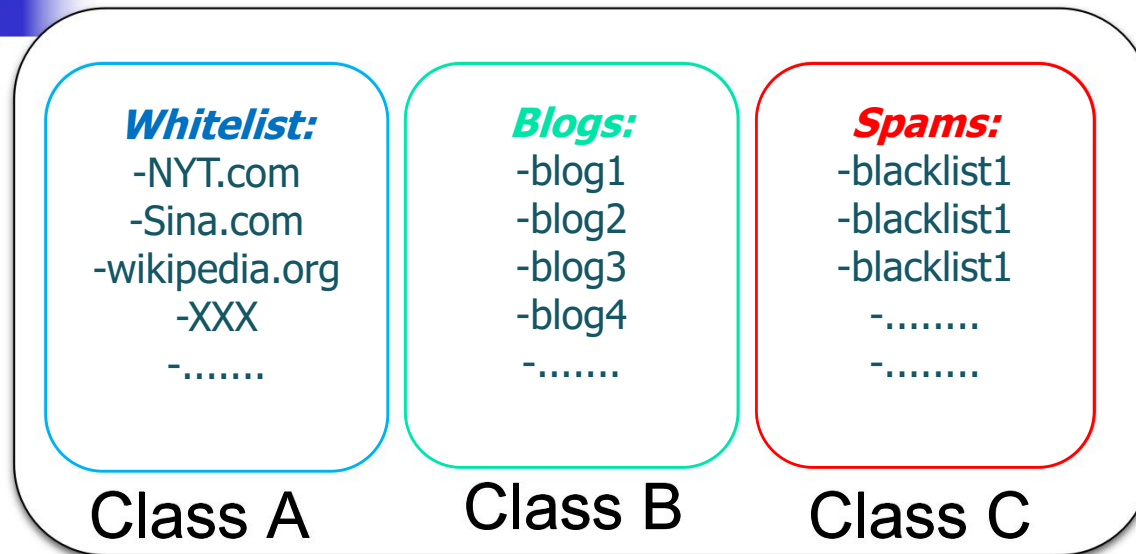


Steps:

- Get datasets: search the web
- 3 Classes:
 - list of **trustworthy URLs** (whitelist) = CLASS 1
 - list of **SPAM URLs** (blacklist) = CLASS 2
 - list of **blogs** = CLASS 3
- Create a CSV file for the data



Dataset:



| Contents | Label |
|---------------------------|-------|
| Content1 (from Whitelist) | A |
| Content2 (from Blogs) | B |
| Content3 (from Spams) | C |



Steps:

- get FINAL DATABASE:

Transform the content (e.g. using TF-IDF)

- Apply the algorithms using scikit-learn (e.g. SVM, NB, etc.)

- Generate the final report:

Plot the graphs and also the confusion matrix