

Undergraduate Program in International Relations

Data Analysis in Politics and Journalism (Winter 2019)

Course objectives:

To provide students with an introduction to quantitative methods in political science and to enable them to analyze information in the media space through objective approaches.

Learning Outcomes:

- Gain basic understanding of text mining
- Develop R programming skills that will enable you to conduct your own research
- Apply data analysis techniques to spot fake news

Course outline

1. Defining Fake News

Misinformation ecosystem. Attributes and classification of fake news. Examples of fake news. Verification tools. Quantifiability of fake news main features.

2. Main sources of data

Parsing data from social networks to R. Parsing data from Twitter. Using publicly available datasets. Data labeling. R packages to analyze unstructured text data. Text data wrangling. Features selection.

3. Logistic regression

Generating features to be used in logistic regression. Estimating parameters. Evaluating quality of the logistic model. Interpretation of the results.

4. Random forests

Decision tree learning. Ensemble methods. Bootstrap aggregating. AdaBoost. Building random forests. Random forests limitations.

5. Naïve Bayes

Introduction to Bayesian methods. Probabilistic model. Deriving classifier from the probability model. Relation to logistic model. Estimating parameters of the classifier.

6. Fake news project

Selecting the best classifier for identifying fake news. Building an automated system for news classification.

Duration: Winter 2019 (Module 3)

Course Materials:

- Instructor's Handouts
- First Draft resources: <https://firstdraftnews.org>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151.
- Wang, W. Y. (2017). "liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648*.
- Atodiresei, C. S., Tănăselea, A., & Iftene, A. (2018). Identifying Fake News and Fake Users on Twitter. *Procedia Computer Science*, 126, 451-461.
- Aldwairi, M., & Alwahedi, A. (2018). Detecting Fake News in Social Media Networks. *Procedia Computer Science*, 141, 215-222.
- Figueira, Á., & Oliveira, L. (2017). The current state of fake news: challenges and opportunities. *Procedia Computer Science*, 121, 817-825.
- Stukal, D., Sanovich, S., Bonneau, R., & Tucker, J. A. (2017). Detecting bots on Russian political Twitter. *Big data*, 5(4), 310-324

Course Structure: The course revolves around emerging approaches to fake news detection, which will be illustrated through case studies and classification algorithms in R.

Forms of Final Assessment: home assignments+group project

Module Grade: 50% - home assignments, 50% - group project

(96-100% - 10, 90-95% - 9, 80-89% - 8, 75-79% - 7, 65-74% - 6, 55-64% - 5, 45-54% - 4, 35-44% - 3, 25-34% - 2, 0-24% - 1)

Instructors: Mikhail Vladimirovich Kamrotov (kamrotov@gmail.com)

Office hours: by appointment

Classroom policies:

- **Hand-in assignments policy:** All home assignments should be submitted electronically via instructor's email on the due date. No deadline extensions are possible.
- **Cheating policy:** In case of any kind of plagiarism (with the detected source), the assignment is evaluated as zero without the chance to make up for it. In case of two

written assignments with the similarity index of 50% and higher from two students, both get a zero for the assignment.