

ACA assignment

Applied Data Science Project

This assessment is the capstone project of the module. It requires you to address a crime and security science research problem in the full data science workflow (e.g., obtaining the data, processing the data, modelling the data, building predictive models, reporting on the findings, interpreting the outcomes). You will write a brief report on your project (a template will be provided) and you have to submit the R code needed to reproduce your findings. After passing this assessment, you will have demonstrated the skills to solve a problem using data science techniques.

Basic info

- Weight for final grade: 70%
- Learning outcomes tested:
 - demonstrating knowledge of a broader range of analytical techniques used in the field of Security and Crime Science
 - performing data science analyses on crime and/or-security related issues
 - applying the data science pipeline on crime and/or-security related issues
 - interpreting and effectively reporting the results of said techniques
- Deadline: **16 April 2019**.
- Feedback deadlines: **8 March 2019** for the peer-feedback, **22 March 2019** for the 1-on-1 feedback (see below)
- Word count limit: 2000 words (excl. code supplement; do not exceed this word count limit!)

Project requirements

You are free to choose a topic that you want to address using data science techniques learned in this module. The problem should be crime/security related.

We strongly encourage ambitious/daring/new/fancy ways of addressing a problem and originality is highly valued in this assignment.

There are few cornerstones for this project that we expect you to include. In this module, we cover three blocks of data science for crime science: (1) obtaining data with webscraping/APIs, (2) using and analysing text data, and (3) applying machine learning models.

In this module you have to use at least 2 of these 3 blocks. The combination of blocks is up to you and you are free to include all three. Some inspirations for each block are listed below (this list is of course not exhaustive):

Block 1: Obtaining data with webscraping/APIs

- retrieving data from APIs that were covered in the module
- new APIs that have a similar R access pipeline
- obtaining data through custom-made webscraping scripts

Block 2: Using and analysing text data

- modelling texts through linguistic representations
- processing + cleaning texts for further analysis
- using text metrics to compare groups/clusters of texts

Block 3: Applying machine learning models

- building predictive models (supervised machine learning)
- applying clustering methods (unsupervised machine learning)

Grading guidelines

Requirements to be graded:

- submit the peer-feedback form in time
- submit the 1-on-1 feedback form in time
- attend the feedback sessions
- submit the report + code supplement before the deadline

Grading criteria:

- Originality of research question/problem – 15%
- Quality of data science techniques – 20%
- Quality of analytical methods – 15%
- Interpretation of findings – 20%
- Clarity of report – 10%
- Quality of report (layout, formatting) – 10%
- Quality of R code (clarity, documentation, reproducibility) – 10%

Feedback sessions

Since a full project is a major step in your data science skills career, we will hold two feedback sessions to help you in the process.

- Peer-feedback session: you will exchange an outline of your project idea (i.e. which problem do you want to address and how?) with a fellow student. The purpose of the peer feedback is to get an independent view on your project early in the process. The peer-feedback session will be held at end of the “Advanced, promises and problems” lecture on **11 March 2019**. Use this template for your feedback submission.
- 1-on-1 feedback session: you will receive individualised feedback from both Bennett and Felix in a 1-on-1 session where we will help you with questions and give you final advice to fine-tune your project. These sessions will take 10 minutes per student and will be held on **25 March 2019** (timeslots to be arranged). Submit your feedback on Moodle through using this template.

Deliverables

- Feedback report for peer-feedback
- Feedback review report for a fellow student
- Feedback report for 1-on-1 session
- Report of your project: the report has a word count limit of 2,000 words and should contain the following sections
 - intro to the problem
 - aim of your project
 - data (i.e. how were the data obtained, which data were used, ...)
 - method (i.e. what did you do with the data, which analytical models did you apply, ...)
 - analytical plan (i.e. clearly set out which analysis you used)
 - results
 - discussion (incl. limitations of your work + outlook on future research)
- Code supplement: you also submit the code needed to fully reproduce your work in an R Notebook
 - the code should be clearly documented
 - the header should contain your student assessment ID (i.e. the code is fully anonymised)

- the code must be reproducible both for the data collection (in case of webscraping) and for the analysis
 - Data: submit your data in the format that you use for the analysis as well as in the raw format
 - raw format: this can be individual text files, existing spreadsheets, news paper articles that were scraped, etc.
 - analysis file: this is typically a `.csv` file or an `.RData` file that you load in when you analyse the data
-