Coursework SECU0057

Data Science for Crime Scientists

Note: this coursework is the replacement of the class test due to the Corona virus situation.

Coursework assessment

- Weight for final grade: 30%
- Learning outcomes tested: (1) demonstrating knowledge of a broader range of analytical techniques used in the field of Security and Crime Science, (2) understanding the purpose, advantages and disadvantages of different forms of data science techniques, (3) interpreting the results of data science techniques.
- Deadline: 30 April 2020, 4pm.
- Word limit: 400 words per question.

Questions

These questions are to be answered in addition to your main data science project. The grading will exclusively depend on the quality and correctness of the methods and not on whether the findings of your analyses are supporting previous work or not.

All questions are to be answered using your dataset that you extracted for your own project.

References are not required for this assessment.

1. Conduct a sentiment trajectory analysis on (at least) five authors in your dataset. How do the average trajectories differ?
2. Use q-gram-based similarity calculations and analyse (for at least two authors) whether the internal similarity (i.e. the similarity among texts of a single authors) is higher than the similarity across authors (i.e. the similarity between texts written by different authors).
3. Extract the overall sentiment of each text in your corpus (i.e. a single score for each text). Now run build and run a supervised classification model and report detailed performance metrics (i.e. accuracy, precision, recall, F1 score, and area-under-the-curve).
4. Choose five texts - each of a different author. For each text find another text in your whole corpus that is most similar in terms of (i) unigram usage, (ii) word length, and (iii) the occurrence of stopwords.

The report

Report this assignment as follows:

- submit a **pdf file** with the filename "SECU0057_coursework_12345.pdf" (replace 12345 with your examination number)
- answer each of the four questions.
- state the word count for each question.
- figures **do not** count towards the word count.
- tables *do* count towards the word count.
- add your R code to the appendix (the appendix is not included in the word count).
- submit the file by the deadline through turn-it-in.

Grading criteria

Criterion	Meaning	Weight
Quality of the answer to Q1	The quality of the solution provided to question 1 (incl. R code).	25%
Quality of the answer to Q2	The quality of the solution provided to question 2 (incl. R code).	25%
Quality of the answer to Q3	The quality of the solution provided to question 3 (incl. R code).	25%
Quality of the answer to Q4	The quality of the solution provided to question 4 (incl. R code).	25%