3,000-words essay (100%, UG) | 4,000 words essay (100%, PG)

In your individual essay you are expected to apply one of the social data science methods covered in this module: web data collection, supervised machine learning, or unsupervised machine learning.

Examples of how to approach the assessment

(1) Based on the R code for Weeks 4 and 5 you can web scrape content/tables from web pages or content from the Guardian and analyse the data using descriptive and bivariate, as well as multivariate analysis, or text analysis in the case of text data.
(2) Based on the R code for Weeks 7 and 8 you can use household survey data for a specific country (e.g., World Values Survey data, or European Values Study, or Pew research data, [Kaggle]) and predict a binary outcome at hand (e.g., health status, opinion on immigration, etc.) using a supervised machine learning algorithm (e.g., decision trees, k nearest neighbours, logistic regression). For this analysis you need to split the data in training and test data.
(3) Based on the R code for Weeks 9 and 10 you can use household survey data for a specific country (e.g., World Values Survey data, or European Values Study, or Pew research data, [Kaggle]) and explore patterns in the data (e.g., latent attitudes related to immigration) using an unsupervised machine learning algorithm (e.g., factor analysis, cluster analysis). If you use the results of this exploratory analysis as an input for prediction, you need to split the data in training and test data.

Indicative structure

Introduction

- Motivate and explain the topic of your analysis (why is it important to study the topic?).
- Clearly state the research aim (e.g., prediction of XYZ in case of supervised machine learning).

Literature Review

- Provide background information about the topic,
- Clearly define relevant concepts,
- If applicable, refer to the theoretical background.
- Present previous research on the topic.

Data and Methods

- Describe the data used (date of data collection, how data was collected, sampling). If you use web-scraped data, describe the web scraping in this section.
- Include a description of all variables considered in the analysis (how each variable is defined and measured).
- Present descriptive statistics (frequency, minimum, maximum, mean and standard deviation, and median for each variable; descriptive figures depend on the type of variable).

Results

- Present your main analysis.
- Depending on the topic, you can also present the descriptive statistics in the section.
- Aim to build up your analysis, starting with a descriptive or bivariate analysis, followed by techniques involving more than two variables.
- When using supervised machine learning the splitting of the data into training and test data should be described in this section. Descriptive figures for both training and test data should be presented.
- If possible, aim to present your findings using figures and/or tables.

Discussion and Conclusions

- Discuss your results in the context of the introduction/literature review and draw main conclusions.
- Discuss the limitations of your research.
- Point to future research that can be conducted.

References

- Use references throughout the essay,
- Use a consistent referencing style such as Harvard Referencing.

Appendix

- Include the R code in the appendix.
- If applicable, present additional analyses in the appendix and refer to it in the main text. All tables and figures presenting the main results should be presented in the main text.