JÖNKÖPING UNIVERSITY
Jönköping International Business School

COURSE SYLLABUS
Methods to Analyze Text as Data, 7.5 credits

Methods to Analyze Text as Data, 7,5 högskolepoäng

Course Code:	FJATD33	**Education Cycle:**	Third-cycle level
Confirmed by:	Jun 16, 2021	**Research subject:**	Economics
Valid From:	Apr 19, 2023		
Version:	1		

Purpose

Human communication is increasingly recorded as digital text, which constitutes big data that can be
used to study numerous scientific and real-world problems. The goals of this course are to (i) provide
an introduction to quantitative methods designed to analyze text, (ii) give an overview over common
applications of these methods in economics and the social sciences, and (iii) illustrate the potential of
text-as-data methods to ask new research questions and find new answers to existing problems.

Intended Learning Outcomes (ILO)

On completion of the course, students will be able to:

Knowledge and understanding
1. Demonstrate knowledge of prominent methods to manage and analyze (large) text, including natural language processing, text classification, sentiment analysis, and topic modeling.
2. Recognize typical research applications using these methods in students' field of study.

Skills and abilities
3. Carry out preliminary steps necessary to analyze text, such as creating a corpus; removing white space, numbers, and stop words; word stemming; and lemmatization.
4. Use existing packages in a programming environment (e.g., Quanteda in R) to describe patterns in text.
5. Write and execute scripts to implement data-as-text methods.
6. Visualize results using word clouds and appropriate diagrams.
7. Exploit online resources (e.g., Stack Overflow) to find solutions for common methodological problems.

Judgement and approach
8. Select the optimal (mix of) method(s) for a given problem.
9. Draw adequate conclusions from text analyses.

Contents

The course provides an overview over the most common text-as-data methods as well as their typical
areas of application:
- Prerequisites (text import, creation of corpora, pre-processing, creation of document-term matrices, lemmatization)
- Text statistics (e.g., frequency analysis, measures of readability, similarity indices)
- Generic and customized dictionaries
- Sentiment analysis
- Text classification using reference texts and supervised learning
- Topic modeling

Type of instruction

The course uses lectures, seminars, tutorials.

The teaching is conducted in English.

Prerequisites

Admitted to a doctoral program in Economics, Business Administration, or a related subject of a recognized business school or university. Knowledge in basic statistics and previous experience with R are recommended, but not required.

Examination and grades

The course is graded Fail (U) or Pass (G).

? Active seminar participation in the form of discussing selected research papers (20%) fulfils ILOs 1 and 2.
? A written assignment in the form of an R script fulfils ILOs 3-9.
The grades are "pass" or "fail". All examination elements need to be completed for a "pass".

Course evaluation

A course evaluation will be conducted at the end of the course.

Other information

Connection to Research and Practice
The course helps its students to understand research designs in economics and related disciplines
(e.g., communications, finance, management, marketing, and political science) that rely on text-asdata
methods. In addition, students can directly apply these methods in their own research. The methods can be used to answer research questions in JIBS' focus areas, such as "What text characteristics determine the success of a crowdfunding campaign?" (entrepreneurship), "Can social
media influencers affect the adoption of green technologies?" (renewal), and "What media slant do

newspaper owners provide?" (ownership). As these examples illustrate, the course draws on contemporary problems in business and society, while providing a sound scientific basis of the relevant methods.

Course literature

Beattie, G. (2020). Advertising and media capture: The case of climate change. Journal of Public Economics, 188.

Garz, M. (2020). Quantitative methods. In M. B. von Rimscha (ed.): Management and Economics of

Communication, pp. 109–128. De Gruyter, Boston, MA.

Garz, M., & Martin, G. (2020). Media Influence on Vote Choices: Unemployment News and Incumbents' Electoral Prospects. American Journal of Political Science, forthcoming.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as Data. Journal of Economic Literature, 57, 535–

574.

Gurun, U. G., & Butler, A. W. (2012). Don't Believe the Hype: Local Media Slant, Local Advertising,

and Firm Value. Journal of Finance, 67, 561–598.

Jetter, M. (2017). The effect of media attention on terrorism. Journal of Public Economics, 153, 32–

48.

Jurafsky, D., & Martin, J.H. (2008). Speech and Language Processing, 2nd edition. Prentice Hall, Upper

Saddle River, NJ.

Lee, D., Hosanagar, K., & Nair, H. S. (2018). Advertising Content and Consumer Engagement on Social

Media: Evidence from Facebook. Management Science, 64, 5105–5131.

Taeuscher, K., Bouncken, R., & Pesch, R. (2020). Gaining Legitimacy by Being Different: Optimal

Distinctiveness in Crowdfunding Platforms. Academy of Management Journal, forthcoming.

All course literature will be provided at the beginning of the course.