UNIVERSITÄT BAYREUTH

Economics VII: Chair of
Quantitative Economic History

Summer Term 2024 **Bayreuth, 28.03.2024**

Data Science in Economic History: Text as Data

Timur Öztürk

Tuesdays, 14:15-15:45, S56 (RW I / PC-Pool)

Course Summary

This course introduces economic history students to data science, focusing on text analysis and natural language processing (NLP) using Python. Students will learn to process, analyze, and visualize text data from various sources, including historical documents, economic reports, and digital archives. The course covers basic programming in Python, text preprocessing, sentiment analysis, topic modeling, and other advanced NLP techniques. By the end, students will complete a project demonstrating their ability to apply text analysis methods to a historical economic research question.

General Information

The course will be taught in **English.**

Target Groups:

- Master's students in History & Economics, Philosophy & Economics, Economics, or related programmes.
- Advanced undergraduate students with a profound interest in econometrics and Python.

Prerequisite: Students should be familiar with basic econometric methods. Prior exposure to Python is recommended (although not strictly necessary) as we will not learn Python from scratch.

Enrolment is restricted to 20 students. Please sign up for the course on CMLife (first come, first serve). Registration opens on 01.03.2024 and closes on 15.04.2024.

We will make slides, papers, and other resources available via e-learning.

Grading

- Two take-home assignments: 25% (12.5% each), due 14.05. and 25.06.2024.
- Final Project: 15% for the Presentation, 60% for the final submission due 30.09.2024.

Literature

The following two online textbooks cover many of the tools and concepts covered in this course:

Prof. Dr. Sebastian T. Braun I Economics VII: Chair of Quantitative Economic History, Universitätsstr. 30, 95447 Bayreuth
Tel: 0921 55-6256, Fax: 0921 55-84 6241, sebastian.braun@uni-bayreuth.de
https://www.econhist.uni-bayreuth.de

- Bird S. Klein E. & Loper E. (2009). *Natural language processing with python*. O'Reilly. (Short: NLP)
- Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press. (Short: TAD)

In addition, we will read and replicate selected journal articles, which I will make available on e-learning. I will also provide more concise resources on e-learning each week, such as useful websites.

Preliminary Schedule

Date	Topic	Reading / Lab Activity / Assignments
16.04.	Introduction to the Course & Python	NLP Chapter 1
23.04.	Regular Expressions & Tokenization	NLP, Chapter 2
30.04.	Pre-processing & Bag of Words	NLP, Chapter 3 & TAD, Chapter 5
07.05.	Application Session I	Own Resources
14.05.	Topic Identification & Classification	Assignment I Due Date
28.05.	Text as Outcome & Treatment	NLP, Chapter 5 & TAD, Chapter 17
4.06.	Sentiment Analysis & Word Clouds	TAD, Chapters 25 & 26
11.06.	Predicting Sentiment & Visualization	NLP, Chapter 10 & Extra Resources
18.06.	Supervised & Deep Learning for NLP	Own Resources
25.06.	Application Session II	Own Resources & Assignment II Due Date
02.07.	Application Session III	Own Resources
09.07	Project Presentations I	TBA
16.07.	Project Presentations II	TBA

Office Hours

I will be available for office hours between 12:30 – 14:00 on Tuesdays.

Academic Integrity

Students are expected to adhere to the University of Bayreuth's standards of academic integrity. I encourage you to collaborate with your peers and use the online tools available to you. Nevertheless, any submitted work must be your own, uniquely.

Prof. Dr. Sebastian T. Braun I Economics VII: Chair of Quantitative Economic History, Universitätsstr. 30, 95447 Bayreuth
Tel: 0921 55-6256, Fax: 0921 55-84 6241, sebastian.braun@uni-bayreuth.de
https://www.econhist.uni-bayreuth.de