UEA Summer University

Text as Data: A Hands-On Introduction to Textual Analysis in Python

Programme Overview	
Programme Title: Text as Data: A Hands-On Introduction to Textual Analysis in Python	
Credit Value: Non-credit bearing	**Programme Level:** MsC and PhD students
Period: Summer (1-week course)	**Programme Dates:** July/August 2026

Programme Description and Learning Outcomes	
Programme Description	This summer school provides a structured introduction to textual analysis for MSc and early PhD students. The course progresses from rudimentary frequency-based approaches to probabilistic models, topic modelling, and modern large language models (LLMs). The unifying theme is that all textual methods involve projecting text into numerical or latent representational spaces, with different assumptions and inferential consequences. The course explicitly compares methods with and without LLMs, emphasising construct validity, measurement error, and researcher discretion within the social sciences and with a focus on business topics.
Summary Overview	**Sunday: Arrival to Norwich and UEA campus** Each morning consists of an expositive lecture. Afternoons are hands-on practical sessions. **Monday:** Foundations and frequency-based approaches In the morning, students will learn about text as data, why textual analysis matters, core dimensions of text, an overview of methods, and Python and computational setup. After a lunch break, the afternoon session will cover Python fundamentals, text ingestion, tokenisation, word counts, and exploratory analysis. **Tuesday:** Dictionaries, scaling, and probabilistic language models The morning session will cover 'Bag-of-words' models, dictionary methods, scaling, probabilistic language models, and assumptions and limitations. The afternoon session will study sentiment analysis, dictionary-based classification, unigram models, and sensitivity to preprocessing. **Wednesday:** Topic modelling This morning session will cover topic modelling, latent structure, LDA, and interpretation versus identification. The afternoon will be free time for social activities.

	Thursday: Embeddings and pre-LLM semantic methods The morning will have presentations on distributional semantics, embeddings, similarity, clustering, and transition from topics to continuous representations. In the afternoon, students will have a presentation on an empirical comparison of topic models and embeddings. **Friday:** Large language models and comparative inference The morning will focus on large language models, transformer intuition, prompting, representational gaps, and construct validity. The last afternoon session will offer a comparative analysis of identical tasks with and without LLMs and a talk on reproducibility, ethics, and inference. **Saturday:** Departure from UEA campus.
Learning Outcomes	Upon successful completion of this course, students will be able to conceptualize textual data as a model-dependent object of empirical analysis; implement and critically evaluate a range of textual analysis methods, including frequency-based approaches, probabilistic language models, topic models, embedding-based representations, and large language models; assess the assumptions, limitations, and sources of measurement error associated with each methodological family; apply identical research questions using both LLM-based and non-LLM-based approaches and systematically compare their outputs; and transparently document methodological choices, interpret results in light of construct validity concerns, and communicate findings in a manner consistent with reproducible and credible social science research.
Core Reading List	Participants will need access to required coding and programming tools. All practical sessions use Python, primarily via Anaconda and Jupyter. Core libraries include pandas, numpy, scikit-learn, gensim, and sentence-transformers.
Programme Assessment	There is no assessment for completion of the course.
Modes of Programme Delivery and Credits	Students will receive a certificate of completion at the end of the course.
Credits	This is not a credited course, but students will receive a certificate of completion.