

MODULE SPECIFICATION • ELECTIVE

Natural Language Processing

The purpose of this course is to introduce students to the theory and practice of applying natural language processing (NLP) in economics and business. The course covers all steps in the data science pipeline of transforming textual data into numbers that are relevant for decision making. It assumes no prior knowledge concerning specific NLP related subjects and starts with a general introduction to text mining. After the introduction, it covers the most used unsupervised methods, such as matrix factorization and topic modelling, and (semi-)supervised methods, such as document classification. The course is also designed to enable students to study the principles of constructing linear econometric time series models and how these models can be used in various practical contexts. There are no prerequisite modules required before taking this course.

10	15h	85h	100h
CREDITS	CONTACT HOURS	PRIVATE STUDY	TOTAL STUDY

COURSE CONVENOR

NAME Dr Matthieu Picault

OFFICE By appointment — please email to arrange a convenient time

COURSE STRUCTURE

Econ 7916 is a 10-credit course and students are expected to input approximately 100 hours of study. The total number of contact hours is 15 hours, leaving 85 hours for private study. Delivery is by lectures — 15 hours across the first three weeks of term, comprising 10 hours of lectures and 5 hours of tutorials. Optional clinics run on the last day of the course.

During private study, balance the course material (the primary source) against the recommended textbooks: think critically about how these fit into the body of knowledge, perform the exercises, complete coursework and revise. You can expect to perform well only by working consistently through the year.

MODULE PROGRAMME

LECT.	TOPIC	SESSION CONTENT
01	Explorative Textual Data Analysis	- Introduction to applications of NLP in economics and management - Text as data - Cleaning textual data - Document feature matrix and word clouds
02	Categorizing documents	- Unsupervised topic models - Supervised document classification
03	Sentiment analysis	- Definition of sentiment - Lexicon-based approaches - Improvements
04	Time series analysis with sentiment	- Sentiment aggregation into time series (Sparse) prediction and forecasting evaluation - Introduction to the assignment

KEY SKILLS

By the end of this course, students should have some knowledge and understanding of:

- Relevant natural language programming techniques and their application to a wide range of empirical issues
- Integration of natural language programming techniques in econometric prediction models
- Use of the open source statistical software Python for solving NLP problems in economics and business

DESIRED OUTCOMES

By the end of this course, students should be able to:

- Engage in abstract thinking by extracting the essential features of large textual corpora to solve economic and managerial problems
- Communicate and present complex arguments in oral and written form with clarity and succinctness
- Present, interpret and analyse information in numerical form
- Utilise effectively statistical and other packages
- Work effectively both individually and within a team environment

ASSESSMENT

WEIGHTING	100% coursework
ISSUED	End of week 6
DUE	End of week 10 — four weeks for completion
SUBMISSION	Electronically via the Moodle site for this course
FEEDBACK	Returned within 4 weeks of the deadline, as marker's notes appended to the PDF of your coursework

Marking criteria and penalties are set out in the Economics Undergraduate Handbook and the general course information paper, available via the Current Student page of the university website under Academic Regulations.



COURSE TEXT & RECOMMENDED READING

Copies of the lecture slides are made available on the course web pages, together with additional material and course announcements.