# Econ 616 Natural Language Processing  2023-2024

## COURSE AIMS & OBJECTIVES, KEY SKILLS AND LEARNING OUTCOMES

**Course Aims & Objectives:** The purpose of this course is to introduce students to the theory and practice of applying natural language processing (NLP) in economics and business. We cover all steps in the data science pipeline of transforming textual data into numbers that are relevant for decision making. We assume no prior knowledge concerning specific NLP related subjects and start off with a general introduction into text mining. After the introduction, we cover the most used unsupervised, such as matrix factorization and topic modelling, and (semi-)supervised, such as document classification, text mining methods. In addition, the course is 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.

**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 econom(etr)ic 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 problem
- 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.

## COURSE STRUCTURE

Econ 616 is a 10 credits course and therefore students are expected to input approximately 100 hours of study into the course. The total number of contact hours on Econ 616 is 15 hours. This leaves 85 hours for private study. Course Delivery comes in the form of Lectures with 15 hours delivered over the first 3 weeks of the term (10 hours of

lectures and 5 hours of tutorials). There will be optional clinics on the last day of the course.

During your private study you should strike a balance between reading the course material (which is the primary source of information) and the recommended textbooks, thinking critically about how these fit in to the body of knowledge on the subject and about how our level of knowledge can be improved, performing exercises, completing coursework and revising for examinations. You can expect to perform well on this course only if you work consistently through the year.

## COURSE CONVENOR
Dr. Matthieu Picault

## LECTURERS CONTACT INFORMATION (Including Office Hours)
Email: matthieu.picault@univ-orleans.fr.

Available by appointment (please email to arrange a convenient time)

## COURSEWORK ASSESSMENT
The CWA mark will be calculated as 100% coursework. The coursework will be assigned at the end of the course.

The coursework will be delivered to students at the end of week 6 (23 February) and is due for submission at the end of week 10 (22 March), allowing students 4 weeks for completion.

Coursework must be submitted electronically through the Moodle site for this course: https://mle.lancs.ac.uk/course. Login using your regular Lancaster University access details. This opens a page headed MLE: My home.

The format of the submission is as follows.

- The submitted file must be in pdf format with the following name

  **stud#_studname_cw_cw#.pdf**

  where: **stud#** is your student number, **studname** is your name in the format *surname_name*, **cw#** is either 1 or 2 according to the piece of coursework submitted. Eg a student with student number 111 would submit a file named *111_surname_firstname_cw_1.pdf*.

- Maximum file size is 2MB: figures resolution must be adjusted accordingly.

Note that your work will be screened using software designed to detect plagiarism.

Do not rely upon someone else to submit your coursework.
*Word counts are inclusive of all material submitted apart from the Bibliography.

## FEEDBACK ON COURSEWORK:

The coursework will be marked and returned to students within 4 weeks of the submission deadline. Feedback will consist of marker's notes appended to the pdf of your coursework.

## MARKING CRITERIA AND PENALTIES

Marking criteria can be found in the Economics Undergraduate Handbook and the general course information paper. An electronic copy of this can be found via the Current Student page of the university website then follow the Academic Regulations link
https://gap.lancs.ac.uk/ASQ/QAE/MARP/Documents/UG-Assess-Regs.pdf

## FINAL MARK INFORMATION

This course is assessed 100% by means of coursework. The final mark is the average of the marks obtained in the two pieces of coursework.

## COURSE TEXT AND RECOMMENDED READING

Copies of the lecture slides and Python code will be made available on the course web pages. You **MUST** print off the notes for each lecture **prior to** attending. Solutions to exercises, and some additional material associated with these lectures and course announcements will also be placed on this website.

## COURSE OUTLINE/LECTURE SCHEDULE

*Lecture 1:*     **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

*Lecture 2:*     **Categorizing documents**

- Unsupervised topic models
- Supervised document classification

*Lecture 3: Sentiment analysis*

- Definition of sentiment
- Lexicon-based approaches
- Improvements

*Lecture 4: Time series analysis with sentiment*

- Sentiment aggregation into time series
- (Sparse) prediction and forecasting evaluation
- Introduction to the assignment