

---

## Text Analysis

Period: a.y. 2025/26 – II sem.

Class times:

**Instructor:**

Prof. Qiaoni Shi  
Dept. of Marketing - Room 4-D1-14  
[qiaoni.shi@unibocconi.it](mailto:qiaoni.shi@unibocconi.it)

---

### Course description

This course is structured to explore the expanding realm of textual analysis, a field gaining prominence through its increasing presence in top Marketing and Management journals. It offers an understanding common and emerging methods of analyzing large collections of textual data.

Students will be introduced to a range of text analytics methodologies, alongside acquiring hands-on experience in implementing these methods using Python. The curriculum emphasizes the essential principles and practices of text analytics, their practical application in Python, and their relevance to empirical research.

### Learning Objectives

Upon completing this seminar, students are expected to have developed a set of key skills and understandings:

- A comprehensive understanding of the application of text analytics in the fields of Marketing and Management.
- Insight into both traditional and emerging methodologies for analyzing textual information.
- Proficiency in the basic syntax of Python programming language.
- The ability to effectively apply text analytics methods using Python.

### Course Requirement and Assessment

#### 1. Class Participation and Exercises

Students are expected to complete in-class exercises, which are aimed at reinforcing lecture concepts. Students are encouraged to actively contribute to class discussions. Key indicators of effective participation include diligent preparation for lessons, consistent and active attendance,

and a proactive approach to the learning process.

## 2. **Final Project**

The final project is an opportunity for students to develop a research idea that applies the principles of text analytics within their field of interest. The objective is to initiate the publication process by encouraging students to discover applications of text analysis in their specific areas. The project should result in an innovative concept potentially leading to a publishable manuscript. This could involve a novel research question, hypothesis, or a fresh approach to an existing question using new data for text analysis.

The project submission must include the following components:

- 1) Research Question: Clearly define the research question.
- 2) Motivation: Explain the significance of the research question.
- 3) Literature Review: Discuss relevant theories and prior published works related to the topic.
- 4) Proposed Research Design: Detail how text analysis will be employed to address the research question.
- 5) Data: Describe the data to be used, how it will be sourced or collected, and assess the feasibility of obtaining this data.
- 6) Proposed Analyses: Outline the planned analysis methods.
- 7) Expected Findings and Managerial Implications: Discuss the potential outcomes and their implications.
- 8) Code: Include the code developed for data collection and/or analysis pertinent to the research question.

## **Topics**

### **Lecture**

1. Applications of text analytics in Marketing, Management
2. Natural Language Processing  
Regular Expression, Stop-words, Tokenization, Normalization, Stemming and Lemming, Part of Speech Tagging, Word vector representation: word2vec
3. Text Analysis  
Frequency analysis, Term weighting, Clustering, Classification, Probabilistic topic model (Latent Dirichlet Allocation)

### **Computer Lab**

1. Python Syntax
2. Data Acquisition from the Web: website scraping
3. Text Analytics in Python

## **Python Software**

Google Colaboratory will be used in this course for the Python programming.

## Tentative schedule

Session	Lecture	Lab/Discussion
1	Introduction	Intro to Python and Google Colaboratory
2	Text analysis basics	Python basics
3	Dictionary-based Text analysis	Data acquisition, Regular expression
4	Text pre-processing	Text pre-processing
5	Text representation, e.g., Word2Vec	Literature Discussion
6	Latent Dirichlet Allocation (LDA)	LDA, Sentiment analysis

## Suggested complementary readings

### Methods

David Blei (2012) Probabilistic Topic Models. *Communications of the ACM* 55(4) 77-84

Natural Language Processing with Python. Available at: <https://www.nltk.org/book/>

### Applications in Marketing and Management

Gabel, S., Guhl, D., & Klapper, D. (2019). P2V-MAP: Mapping market structures for large retail assortments. *Journal of Marketing Research*, 56(4), 557-580.

Fanglin Chen, Xiao Liu, Davide Proserpio, and Isamar Troncoso Cortez (2021). "Product2Vec: Understanding Product-level Competition Using Representation Learning," Working Paper. Available at: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3519358](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3519358)

Wei Guo, Tieying Yu, Javier Gimeno (2017) Language and Competition: Communication Vagueness, Interpretation Difficulties, and Market Entry. *Academy of Management Journal* 60(6): 2073-2098.

Rao, Anita (2022) Deceptive claims using fake news advertising: The impact on consumers. *Journal of Marketing Research* 59 (3): 534-554.

Daniel Blaseg, Christian Schulze, Bernd Skiera (2020) Consumer Protection on Kickstarter. *Marketing Science* 39(1):211-233.

Schoenmueller, V., Netzer, O., & Stahl, F. (2020). The Polarity of Online Reviews: Prevalence, Drivers and Implications. *Journal of Marketing Research*, 57(5), 853-877. <https://doi.org/10.1177/0022243720941832>

Giavazzi, Francesco, Felix Iglhaut, Giacomo Lemoli, and Gaia Rubera. "Terrorist Attacks, Cultural Incidents, and the Vote for Radical Parties: Analyzing Text from Twitter." *American Journal of Political Science* (2023).

