Thematic Course

Academic Year	2023-24
Subject	Statistical methods for text mining with R
Instructor	Alessandro Albano and Mariangela Sciandra
Course description	The Thematic Course on "Statistical methods for text mining with R" provides an overview of statistical techniques for text mining and their applications in various fields. Students will learn how to analyze and extract insights from unstructured text data using the R programming language. The course is divided into two modules. **First module (Mariangela Sciandra):** This module will consist of 4 hours. Topics covered include: introduction to the basics of text mining, text preprocessing, regular expressions and text visualization. **Second module (Alessandro Albano):** This module will consist of 8 hours. Topics covered include: sentiment analysis, topic modeling, and text classification. Through a combination of theory classes and computer labs, students will develop practical skills in text mining and be able to apply these techniques to their own research projects.
Learning Objectives	•Understand the fundamentals of text mining and its relevance in analyzing textual data. •Acquire knowledge of various techniques and algorithms used in text mining. •Develop skills in applying text mining methods using R programming language. •Learn how to preprocess, analyze, and visualize textual data. •Gain insights into the practical applications of text mining in economic and statistical research.
Suggested readings	Julia Silge and David Robinson (2017). Text Mining with R by Julia Silge and David Robinson (O'Reilly), 978-1-491-98165-8." https://www.tidytextmining.com/index.html Kwartler, T. (2017). Text mining in practice with R. John Wiley & Sons. Emil Hvitfeldt and Julia Silge (2022). Supervised Machine Learning for Text Analysis in R. https://smltar.com/
Course Activity (hrs)	12
Credits	3
Assessment Method	Report/presentation on a scientific paper and/or a dataset as agreed with the instructor.
Teaching Methods	Theory classes and computer labs
Calendar	From 26th of February to 1st of March, 2024 (to be agreed with students)

Contacts	alessandro.albano@unipa.it
	mariangela.sciandra@unipa.it