

Syllabus

Course description

Course title	Big data methods for economics and business M1 Statistical Methods for high-dimensional data M2 Natural language processing and web analytics
Course code	27512
Scientific sector	SECS-S/01 + INF/01
Degree	Master in Data Analytics for Economics and Management
Semester and academic year	tbd semester a.y. 2024/2025
Year	2nd study year
Credits	12 (6+6)
Modular	Yes

Total lecturing hours	72 (36+36)
Total lab hours	/
Total exercise hours	/
Attendance	suggested, but not required
Prerequisites	none
Course page	https://www.unibz.it/en/faculties/economics-management/master-data-analytics-economics-management/

Specific educational objectives	<p>This is an advanced course focusing on the analysis of high-dimensional and big data commonly encountered in economics and business. The students will be introduced to modern statistical learning methods drawing together theory, data, computation and recent research. Much emphasis will be given to applications of methods in the domains of economics and business. The first module focuses on theory, statistical models and methods for model selection and aggregation, in the context of various models including regression, latent variable and graphical models. The second module covers the application of algorithms and statistical techniques in the areas of natural language processing and predictive web analytics.</p>
--	--

Module 1	M1 Statistical Methods for high-dimensional data
Lecturer	Greta Goracci
Scientific sector of the lecturer	SECS-S/01
Teaching language	English
Office hours	please refer to the lecturer's timetable
Lecturing assistant	None
Teaching assistant	None

List of topics covered	<ul style="list-style-type: none"> • High-dimensional data, big data and the curse of dimensionality • Convex critehods • Model aggregatation and model combining • Introduction to data dimension reduction • High-dimensional regression • Graphical models • Multiple testing
Teaching format	Frontal lectures, exercises, computer labs, face-to-face discussions and flipped-classroom activities.

Module 2	M2 Natural language processing and web analytics
Lecturer	TBA
Scientific sector of the lecturer	TBA
Teaching language	English
Office hours	please refer to the lecturer's timetable
Lecturing assistant	TBA
Teaching assistant	None
Office hours	please refer to the lecturer's timetable
List of topics covered	<ul style="list-style-type: none"> • Basics of natural language Processing (NLP) • Text classification and prediction using supervised approaches • Unsupervised methods for NLP and latent models • Neural networks for NLP and neural language models • Information Retrieval • Relation extraction, question answering, dialog systems and chatbots • Web crawling and link analysis
Teaching format	Frontal lectures, exercises, computer labs, face-to-face discussions and flipped-classroom activities.

Learning outcomes	<p><u>1) Knowledge and understanding:</u> The student acquires knowledge of analytical tools and statistical techniques needed to understand and analyze high-dimensional and big data frequently used to understand economic and business phenomena. The student acquires in-depth knowledge of advanced statistical methods by developing programming skills and focusing on applications to economic and business data.</p> <p><u>2) Applying Knowledge and understanding</u></p>
--------------------------	--

	<p>The student acquires the ability to apply and implement quantitative methods focusing on different types of data, including text and network data data. These capabilities are declined in various application domains of interest to companies and public and private organizations.</p> <p><u>3) Making judgments</u> Ability to choose appropriate statistical methods to address a given research question and to make effective decisions in support of economic and business issues.</p> <p><u>4) Communication skills</u> Ability to communicate effectively the results from statistical analyses of observed data.</p> <p><u>5) Learning skills</u> The course is aimed to provide the methodological and applied knowledge of statistical methods for high-dimensional and textual data, necessary to address subsequent studies, including advanced courses in mathematics, statistics, computer science, as well as applied projects in laboratories and internships, and empirical analyses in the final thesis.</p>
--	--

<p>Assessment</p>	<p>M1: Final Exam (60%): The final exam consists of problems related to the use a of statistical methods and interpretation of results obtained from the analysis and interpretations of various data sets. Assignments (40%): Data analysis assignments to be handed in will be assigned three times during the semester.</p> <p>M2: Final Exam (60%): The final exam consists of problems related to the use a of statistical methods and interpretation of results obtained from the analysis and interpretations of various data sets. Assignments (40%): Data analysis assignments to be handed in will be assigned three times during the semester.</p> <p>The final exam in each module tests Skill 1 (Knowledge and understanding). The project assignments are computer-based and allow to verify Skills 2, 3 and 4 (Applying knowledge and understanding, Making</p>
--------------------------	--

	judgements, Communication skills). The skill concerned with autonomous study (Skill 5, Learning skills) is indirectly verified, because passing the final exam requires autonomous execution of exercises suggested by the lecturer as well as individual preparation to the class discussions and flipped-classroom activities.
Assessment language	English
Evaluation criteria and criteria for awarding marks	<p>In both modules the exam modalities are the same for both the attending and the non-attending students. Project work (40% of the final grade) and written exam (60% of the final grade).</p> <ul style="list-style-type: none"> • Relevant for project work: clarity of presentation, ability to gain useful and novel insights from data, creativity, critical thinking, ability to adhere to reproducible research best practices • Ability to use R and other software to perform basic data preparation tasks, ability to properly use R libraries, ability to choose the best type of graphical representation for different types of data, correct usage of basic statistical tools • Ability to use Python to employ (understand, recall and use) data analytics methods in practical settings in relation to data analysis and visualization.

Required readings	<p>Lederer, J. (2022). Fundamentals of high-dimensional statistics. Springer International Publishing.</p> <p>Hvitfeldt, Emil, and Julia Silge. Supervised machine learning for text analysis in R. CRC Press, 2021.</p> <p>Jurafsky, David, and James H. Martin. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Upper Saddle River, NJ: Prentice-Hall, 2000. ISBN: 0130950696.</p>
Supplementary readings	TBA