

## Syllabus

### Course description

<b>Course title</b>	<b>Big data methods for economics and business</b> M1 Statistical Methods for high-dimensional data M2 Natural language processing and web analytics – offered online
<b>Course code</b>	27512
<b>Scientific sector</b>	SECS-S/01 + INF/01
<b>Degree</b>	Master in Data Analytics for Economics and Management
<b>Semester and academic year</b>	1 <sup>st</sup> and 2 <sup>nd</sup> semester a.y. 2024/2025
<b>Year</b>	2nd study year
<b>Credits</b>	12 (6+6)
<b>Modular</b>	Yes

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

<b>Specific educational objectives</b>	<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>
--	--

<b>Module 1</b>	M1 Statistical Methods for high-dimensional data
<b>Lecturer</b>	Davide Ferrari <a href="https://www.unibz.it/en/faculties/economics-management/master-data-analytics-economics-management/">Davide Ferrari / Libera Università di Bolzano (unibz.it)</a>

<b>Scientific sector of the lecturer</b>	SECS-S/01
<b>Teaching language</b>	English
<b>Office hours</b>	Please refer to the lecturer's timetable
<b>Lecturing assistant</b>	None
<b>Teaching assistant</b>	None
<b>List of topics covered</b>	<ul style="list-style-type: none"> <li>• High-dimensional data, big data and the curse of dimensionality</li> <li>• Convex critehods</li> <li>• Model aggregation and model combining</li> <li>• Introduction to data dimension reduction</li> <li>• High-dimensional regression</li> <li>• Graphical models</li> <li>• Multiple testing</li> </ul>
<b>Teaching format</b>	Frontal lectures, exercises, computer labs, face-to-face discussions and flipped-classroom activities.

<b>Module 2</b>	M2 Natural language processing and web analytics
<b>Lecturer</b>	Prof. Dr. Paul Pronobis (18 h) <a href="http://unibz.it">Paul Pronobis / Libera Università di Bolzano (unibz.it)</a> External Lecturer TBA (18 h)
<b>Scientific sector of the lecturer</b>	SECS-P/07
<b>Teaching language</b>	English
<b>Office hours</b>	Please refer to the lecturer's timetable
<b>Lecturing assistant</b>	None
<b>Teaching assistant</b>	None
<b>Office hours</b>	Please refer to the lecturer's timetable
<b>List of topics covered</b>	<ol style="list-style-type: none"> <li>1. <b>Introduction to Natural Language Processing (NLP):</b> Exploring the fundamentals of NLP, including its history, applications, and difference to other neural networks.</li> <li>2. <b>Algorithmic Text Classification and Sentiment Analysis:</b> Detailed instruction on various algorithms for categorizing text and extracting sentiment, comparing their effectiveness and use cases.</li> <li>3. <b>Neural Networks in NLP and Language Modeling:</b> An in-depth look at how neural networks are applied in NLP, focusing on using and evaluating different NLP models.</li> <li>4. <b>Advanced Techniques in Information Retrieval:</b> Utilization of cutting-edge neural network strategies combined with vector space</li> </ol>

	<p>models to efficiently retrieve information.</p> <p>5. <b>Web Scraping for Knowledge Construction:</b> Techniques for extracting information from the web to build databases for applications that demand current or extensive factual data.</p> <p>6. <b>Prompt Engineering for Enhanced Language Understanding:</b> Crafting effective prompts to improve relation extraction, answer questions accurately, support dialog systems, and create responsive chatbots.</p> <p>7. <b>Innovations in Large Language Model (LLM) Applications:</b> Exploring multi-agent conversations and the latest advancements in LLM applications, pushing the boundaries of interactive AI systems.</p>
<b>Teaching format</b>	Frontal lectures, exercises, computer labs, face-to-face discussions and flipped-classroom activities.

<b>Learning outcomes</b>	<p>1) <u>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>2) <u>Applying Knowledge and understanding</u> The student acquires the ability to apply and implement quantitative methods focusing on different types of data, including text and network data. These capabilities are declined in various application domains of interest to companies and public and private organizations.</p> <p>3) <u>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>4) <u>Communication skills</u> Ability to communicate effectively the results from statistical analyses of observed data.</p> <p>5) <u>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</p>
--------------------------	--

	<p>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><b>Assessment</b></p>	<p><b>M1:</b>  Final Exam (60%):  The final exam consists of problems related to the use 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><b>M2:</b>  Final Exam (60%):  The final exam consists of problems related to the use 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.</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 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.</p>
<p><b>Assesment language</b></p>	<p>English</p>
<p><b>Evaluation criteria and criteria for awarding marks</b></p>	<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"> <li>• 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</li> <li>• 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</li> </ul>

	<p>representation for different types of data, correct usage of basic statistical tools</p> <p>Ability to use Python to employ (understand, recall and use) data analytics methods in practical settings in relation to data analysis and visualization</p>
<p><b>Required readings</b></p>	<p>Lederer, J. (2022). Fundamentals of high-dimensional statistics. Springer International Publishing.</p> <p>Tunstall, L., Von Werra, L., &amp; Wolf, T. (2022). <i>Natural language processing with transformers</i>. " O'Reilly Media, Inc.</p>
<p><b>Supplementary readings</b></p>	<p>Lecture slides and other reading materials will be provided during the lecture period.</p>