Syllabus

Course Description

Course Title	Big data methods for economics and business
Course Code	27512
Course Title Additional	
Scientific-Disciplinary Sector	NN
Language	English
Degree Course	Master in Data Analytics for Economics and Management
Other Degree Courses (Loaned)	
Lecturers	Prof. Davide Ferrari, Davide.Ferrari2@unibz.it https://www.unibz.it/en/faculties/economics-management/academic-staff/person/39001 Prof. Dr. Paul Michael Pronobis, Paul.Pronobis@unibz.it https://www.unibz.it/en/faculties/economics-management/academic-staff/person/49449
Teaching Assistant	
Semester	All semesters
Course Year/s	2
CP	12
Teaching Hours	M1: - 24 hours of in-person lectures - 12 hours of video lectures (counted as 24 hours to account for re-watching) M2: - 24 hours of in-person lectures - 12 hours of video lectures (counted as 24 hours to account for re-watching)
Lab Hours	-
Individual Study Hours	-
Planned Office Hours	M1: 18 hours

	M2: 18 hours
Contents Summary	Module 1 focuses on advanced statistical techniques for analyzing high-dimensional datasets frequently encountered in business intelligence and economic research. Key topics include penalized and convex optimization methods for model selection (such as LASSO), model aggregation techniques, dimension reduction, high-dimensional regression models, and network-based inference using graphical models. The module also introduces multiple testing procedures for identifying significant patterns across many variables. Emphasis is placed on practical implementation using R and Python, and on the ability to apply these tools to extract interpretable, actionable insights from large-scale data in business and economic applications. Module 2 provides an in-depth introduction to Natural Language Processing (NLP) with a strong focus on modern applications in business and economics. Core topics include algorithmic text classification, sentiment analysis, neural language modeling, and advanced information retrieval using vector-based and neural approaches. Students will learn techniques for web scraping, prompt engineering, and the use of Retrieval-Augmented Generation (RAG) systems, which combine document retrieval with generative models to improve accuracy and relevance. The module also explores recent developments in large language model (LLM) applications, including multi-agent systems and conversational AI, equipping students to critically evaluate and implement state-of-the-art NLP solutions.
Course Topics	M1: • High-dimensional data, big data and the curse of dimensionality • Convex criterions for model selection • Model aggregation and model combining • Introduction to data dimension reduction • High-dimensional regression • Graphical models • Multiple testing M2: 1. Introduction to Natural Language Processing (NLP): Exploring the fundamentals of NLP, including its history, applications, and

	difference to other neural networks.
	2. Algorithmic Text Classification and Sentiment Analysis: Detailed instruction on various algorithms for categorizing text and extracting sentiment, comparing their effectiveness and use cases.
	3. Neural Networks in NLP and Language Modeling: An in-depth look at how neural networks are applied in NLP, focusing on using and evaluating different NLP models.
	4. Advanced Techniques in Information Retrieval: Utilization of cutting-edge neural network strategies combined with vector space models to efficiently retrieve information.
	5. Web Scraping for Knowledge Construction: Techniques for extracting information from the web to build databases for applications that demand current or extensive factual data.
	6. Prompt Engineering for Enhanced Language Understanding: Crafting effective prompts to improve relation extraction, answer questions accurately, support dialog systems, and create responsive chatbots.
	7. Fine-Tuning: Introducing key steps for adapting pre-trained language models (CLM and MLM) through preprocessing and model training. Also covers performance evaluation using tools like Wandb, enabling effective monitoring and optimization for various NLP tasks.
	8. Innovations in Large Language Model (LLM) Applications: Exploring multi-agent conversations and the latest advancements in LLM applications, pushing the boundaries of interactive AI systems.
Keywords	
Recommended Prerequisites	
Propaedeutic Courses	
Teaching Format	The course adopts a blended, student-centred approach that emphasises problem-based learning and active engagement. A portion of the lecture content is made available online in advance, allowing students to explore key concepts independently and at their own pace before attending class. This preparatory work enables in-person sessions to focus on the application of knowledge through real-world problems, collaborative activities, and guided discussions — fostering critical thinking and deeper learning. The course is fully aligned with the principles of the Italian Universities Digital Hub (EDUNEXT) initiative

	(https://edunext.eu), which promotes the integration of digital resources and active learning strategies within university teaching.
Mandatory Attendance	Recommended, but not required.
Specific Educational Objectives and Learning Outcomes	Knowledge and understanding: The student will acquire knowledge of the analytical techniques and tools required to understand and quantitatively analyse economic and business phenomena in order to support decision-making processes. Knowledge of statistical inference, linear models and their generalisations, linear algebra, and optimisation techniques will be consolidated. In-depth knowledge of the main techniques of supervised and unsupervised statistical learning will be acquired, which are functional for the development of analysis and visualisation capabilities of economic and business data. Applying knowledge and understanding: Ability to apply and implement analysis techniques focusing on different types of datasets such as streaming data, tabular data, documents and images and analysis on joint datasets. Ability to apply supervised and unsupervised learning themes, and knowledge modelling, extraction, integration, analysis and exploitation; these skills are declined in various application domains of interest to companies and public and private entities Making judgements: Master graduates will have the ability to apply the acquired knowledge to interpret data in order to make managerial and operational decisions in a business context. Master's graduates will have the ability to apply the acquired knowledge to support processes related to production, management and risk promotion activities and investment choices through the organisation, analysis and interpretation of complex databases. Communication skills: Master's graduates will be able to communicate effectively in oral and written form the specialised contents of the individual disciplines, using different registers, depending on the recipients and the communicative and didactic purposes, and to evaluate the formative effects of their communication.

	Learning skills: "MSc graduates should be familiar with the tools of scientific research. They will also be able to make autonomous use of information technologies to carry out bibliographic research and investigations both for their own training and for further education. In addition, through the curricular teaching and the activities related to the preparation of the final thesis, they will be able to acquire the ability - to identify thematic links and to establish relationships between methods of analysis and application contexts; - to frame a new problem in a systematic manner and to implement appropriate analysis solutions; - to formulate general statistical-econometric models from the phenomena studied.
Specific Educational Objectives and Learning Outcomes (additional info.)	
Assessment	The overall exam mark will be determined by the assessment of the two modules (M1+M2). M1: 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. M2: 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.
Evaluation Criteria	In both modules the exam modalities are the same for both the attending and the non-attending students.

<table>
<tr><td rowspan="2"></td><td>Project work (40% of the final grade) and written exam (60% of the final grade).
•   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.</td></tr>
</table>

Required Readings	M1: Lederer, J. (2022). Fundamentals of high-dimensional statistics. Springer International Publishing. M2: Tunstall, L., Von Werra, L., & Wolf, T. (2022). *Natural language processing with transformers.* " O'Reilly Media, Inc.
Supplementary Readings	
Further Information	
Sustainable Development Goals (SDGs)	

Course Module

Course Constituent Title	M1 - Statistical methods for high-dimensional data
Course Code	27512A
Scientific-Disciplinary Sector	SECS-S/01
Language	English
Lecturers	Prof. Davide Ferrari, Davide.Ferrari2@unibz.it https://www.unibz.it/en/faculties/economics-management/academic-staff/person/39001
Teaching Assistant	
Semester	First semester

CP	6
Responsible Lecturer	
Teaching Hours	- 24 hours of in-person lectures - 12 hours of video lectures (counted as 24 hours to account for re-watching)
Lab Hours	-
Individual Study Hours	-
Planned Office Hours	18
Contents Summary	This module focuses on advanced statistical techniques for analyzing high-dimensional datasets frequently encountered in business intelligence and economic research. Key topics include penalized and convex optimization methods for model selection (such as LASSO), model aggregation techniques, dimension reduction, high-dimensional regression models, and network-based inference using graphical models. The module also introduces multiple testing procedures for identifying significant patterns across many variables. Emphasis is placed on practical implementation using R and Python, and on the ability to apply these tools to extract interpretable, actionable insights from large-scale data in business and economic applications.
Course Topics	• High-dimensional data, big data and the curse of dimensionality • Convex criterions for model selection • Model aggregation and model combining • Introduction to data dimension reduction • High-dimensional regression • Graphical models • Multiple testing
Teaching Format	This module adopts a blended, student-centred approach that emphasises problem-based learning and active engagement. A portion of the lecture content is made available online in advance, allowing students to explore key concepts independently and at their own pace before attending class. This preparatory work enables in-person sessions to focus on the application of knowledge through real-world problems, collaborative activities, and guided discussions — fostering critical thinking and deeper learning. The course is fully aligned with the principles of the Italian Universities Digital Hub (EDUNEXT) initiative (https://edunext.eu), which promotes the integration of digital

	resources and active learning strategies within university teaching.
Required Readings	Lederer, J. (2022). Fundamentals of high-dimensional statistics. Springer International Publishing.
Supplementary Readings	

Course Module

Course Constituent Title	M2 - Natural language processing and web analytics
Course Code	27512B
Scientific-Disciplinary Sector	INF/01
Language	English
Lecturers	Prof. Dr. Paul Michael Pronobis, Paul.Pronobis@unibz.it https://www.unibz.it/en/faculties/economics-management/academic-staff/person/49449
Teaching Assistant	
Semester	Second semester
CP	6
Responsible Lecturer	
Teaching Hours	- 24 hours of in-person lectures - 12 hours of video lectures (counted as 24 hours to account for re-watching)
Lab Hours	-
Individual Study Hours	-
Planned Office Hours	18
Contents Summary	This module provides an in-depth introduction to Natural Language Processing (NLP) with a strong focus on modern applications in business and economics. Core topics include algorithmic text classification, sentiment analysis, neural language modeling, and advanced information retrieval using vector-based and neural approaches. Students will learn techniques for web scraping, prompt engineering, and the use of Retrieval-Augmented Generation (RAG) systems, which combine document retrieval with generative models to improve accuracy and relevance. The module

	also explores recent developments in large language model (LLM) applications, including multi-agent systems and conversational AI, equipping students to critically evaluate and implement state-of-the-art NLP solutions.
Course Topics	1. Introduction to Natural Language Processing (NLP): Exploring the fundamentals of NLP, including its history, applications, and difference to other neural networks. 2. Algorithmic Text Classification and Sentiment Analysis: Detailed instruction on various algorithms for categorizing text and extracting sentiment, comparing their effectiveness and use cases. 3. Neural Networks in NLP and Language Modeling: An in-depth look at how neural networks are applied in NLP, focusing on using and evaluating different NLP models. 4. Advanced Techniques in Information Retrieval: Utilization of cutting-edge neural network strategies combined with vector space models to efficiently retrieve information. 5. Web Scraping for Knowledge Construction: Techniques for extracting information from the web to build databases for applications that demand current or extensive factual data. 6. Prompt Engineering for Enhanced Language Understanding: Crafting effective prompts to improve relation extraction, answer questions accurately, support dialog systems, and create responsive chatbots. 7. Fine-Tuning: Introducing key steps for adapting pre-trained language models (CLM and MLM) through preprocessing and model training. Also covers performance evaluation using tools like Wandb, enabling effective monitoring and optimization for various NLP tasks. 8. Innovations in Large Language Model (LLM) Applications: Exploring multi-agent conversations and the latest advancements in LLM applications, pushing the boundaries of interactive AI systems.
Teaching Format	The module adopts a blended, student-centred approach that emphasises problem-based learning and active engagement. A portion of the lecture content is made available online in advance, allowing students to explore key concepts independently and at their own pace before attending class. This preparatory work enables in-person sessions to focus on the application of knowledge through real-world problems, collaborative activities,

	and guided discussions — fostering critical thinking and deeper learning. The course is fully aligned with the principles of the Italian Universities Digital Hub (EDUNEXT) initiative (https://edunext.eu), which promotes the integration of digital resources and active learning strategies within university teaching.
Required Readings	Tunstall, L., Von Werra, L., & Wolf, T. (2022). *Natural language processing with transformers*. " O'Reilly Media, Inc.
Supplementary Readings	