Syllabus

Descrizione corso

Titolo insegnamento	Big data methods for economics and business
Codice insegnamento	27512
Titolo aggiuntivo	
Settore Scientifico-Disciplinare	NN
Lingua	Inglese
Corso di Studio	Corso di laurea magistrale in Data Analytics for Economics and Management
Altri Corsi di Studio (mutuati)	
Docenti	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
Assistente	
Semestre	Tutti i semestri
Anno/i di corso	2
CFU	12
Ore didattica frontale	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)
Ore di laboratorio	-

Ore di studio individuale	-
Ore di ricevimento previste	M1: 18, including 6 hours offered as virtual classroom sessions M2: 18, including 6 hours offered as virtual classroom sessions
Sintesi contenuti	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.
Argomenti dell'insegnamento	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.
Parole chiave	Inference, big data, high-dimensional, natural language processes, large language models
Prerequisiti	
Insegnamenti propedeutici	
Modalità di insegnamento	Recorded lectures, in-person teaching, exercises. 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.
Obbligo di frequenza	Attendance is recommended, but not mandatory.
Obiettivi formativi specifici e risultati di apprendimento attesi	Intended Learning Outcomes (ILO) M1 ILO 1 Knowledge and understanding: ILO 1.1 The student acquires knowledge of the analytical techniques and tools required to understand and quantitatively analyse economic and business phenomena in order to support decision-making processes. ILO 1.2 The student consolidates knowledge of statistical inference, linear models and their generalisations, linear algebra, and optimisation techniques. ILO 1.3 The student acquires an in-depth knowledge of the main techniques of supervised and unsupervised statistical learning, which are instrumental in the development of analysis and visualisation of economic and business data. ILO 2 Applying knowledge and understanding: ILO 2.1 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. ILO 2.2 Ability to apply supervised and unsupervised learning, and knowledge modelling, extraction, integration, analysis and exploitation; these skills are declined in various application domains of interest to companies and public and private organisations.

	ILO 3 Making judgements: ILO 3.1 The student acquires the ability to apply acquired knowledge to interpret data in order to make directional and operational decisions in a business context. ILO 3.2 The student acquires the ability to apply 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. ILO4 Communication skills: ILO 4.1 The student acquires the ability to communicate effectively in oral and written form the specialised content of the individual disciplines, using different registers, depending on the recipients and the communicative and didactic purposes, and to evaluate the formative effects of his/her communication. ILO 5 Learning skills: ILO 5.1 The student acquires knowledge of scientific research tools. He/she will also be able to make autonomous use of information technology to carry out bibliographic research and investigations both for his/her own training and for further education. Furthermore, through the curricular teaching and the activities related to the preparation of the final thesis, she will be able to acquire the ability - to identify thematic connections 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. M2 ILO 1 Knowledge and understanding: ILO 1.1 The student acquires programming knowledge, particularly aimed

	at data analysis and statistical methodologies for implementing models as well as analysing large-scale datasets. In particular, the computing skills are focused on machine learning methods, on understanding modern techniques for data management and storage, including data from heterogeneous sources in terms of type and structure, such as spatio-temporal data and high-dimensional data, also in cloud environments, and on implementing algorithms for massive data processing. ILO 1.2 Students will acquire knowledge and skills in the analysis of textual data and network structures, with particular attention to issues related to data security and privacy. ILO 2 Applying knowledge and understanding: ILO 2.1 Students will develop the ability to apply and implement techniques for analysing large-scale datasets and spatio-temporal data under conditions of uncertainty, through the design and development of algorithms. The goal is to ensure the utility, quality, and effectiveness of the analysis. ILO 2.2 Ability to use IT technologies, techniques and methodologies for the acquisition, management, integration, analysis and visualisation of large datasets, in order to ensure scalability in terms of dataset volume and acquisition speed. These skills relate in particular to large database and dataset management systems and related visualisation techniques, models and languages for expressing data semantics, learning techniques, decision-making models, information systems organisation, web search techniques and data flow management techniques. ILO 3 Making judgements: ILO 3.1 The student acquires the ability to apply acquired knowledge to interpret data in order to make directional and operational decisions in a business context. ILO 3.2 The student acquires the ability to apply 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. ILO4 Communication skills: ILO 4.1 The student acquires the ability to communicate effectively in oral and written form the specialised content of the individual disciplines, using different registers, depending on the recipients and the communicative and didactic purposes, and to evaluate the formative effects of his/her communication. ILO 5 Learning skills: ILO 5.1 The student acquires knowledge of scientific research tools. He/she will also be able to make autonomous use of information technology to carry out bibliographic research and investigations both for his/her own training and for further education. Furthermore, through the curricular teaching and the activities related to the preparation of the final thesis, she will be able to acquire the ability - to identify thematic connections 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.
Obiettivi formativi specifici e risultati di apprendimento attesi (ulteriori info.)	INTENDED LEARNING OUTCOMES (ILO) ILO 1 – Knowledge and understanding ILO 1.1 The student acquires knowledge and understanding of high-dimensional data and big data settings, including the curse of dimensionality and its implications for statistical modelling and inference. ILO 1.2 The student acquires knowledge and understanding of convex criteria for model selection, model aggregation and model combination methods in high-dimensional contexts.

	ILO 1.3 The student acquires knowledge and understanding of dimension reduction techniques, high-dimensional regression models, graphical models and multiple testing procedures. ILO 1.4 The student acquires knowledge and understanding of the foundations of Natural Language Processing (NLP), including text representation, algorithmic text classification, sentiment analysis and information retrieval. ILO 1.5 The student acquires knowledge and understanding of neural network approaches to NLP and language modelling, including the principles underlying large language models (LLMs) and their evaluation. ILO 1.6 The student acquires knowledge and understanding of advanced NLP applications, including web-based data acquisition, prompt engineering, fine-tuning of pre-trained language models and emerging LLM-based systems. ILO 2 – Applying knowledge and understanding ILO 2.1 The student is able to analyse high-dimensional datasets, identify modelling challenges related to dimensionality, sparsity and multiple testing, and select appropriate statistical tools. ILO 2.2 The student is able to apply convex model selection criteria, aggregation methods, dimension reduction techniques, high-dimensional regression and graphical models to complex data. ILO 2.3 The student is able to implement NLP pipelines for text classification, sentiment analysis and information retrieval, and to evaluate their performance.

	ILO 2.4 The student is able to use neural network–based NLP models, including pre-trained language models, and apply prompt engineering and fine-tuning techniques for specific tasks. ILO 2.5 The student is able to acquire and preprocess textual data from web sources and integrate it into data analysis and NLP workflows. ILO 3 – Making judgements ILO 3.1 The student is able to critically evaluate modelling choices and results in both high-dimensional statistical analysis and NLP applications, taking into account assumptions, uncertainty and performance metrics. ILO 3.2 The student is able to compare alternative statistical and NLP methods and select appropriate approaches based on data characteristics, objectives and computational constraints. ILO 3.3 The student is able to use quantitative and algorithmic evidence to support analytical and operational decisions in data-driven applications. ILO 4 – Communication skills ILO 4.1 The student is able to communicate clearly and effectively, in oral and written form, the methods, results and limitations of high-dimensional statistical analyses and NLP systems, adapting the level of technical detail to different audiences. ILO 5 – Learning skills

	ILO 5.1 The student is able to autonomously deepen knowledge of high-dimensional statistical methods and NLP techniques, integrate new tools and methodologies, and systematically approach new and complex data analysis problems.
Modalità di esame	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 (ILOs 1, 2, 3 , 4 and 5). Assignments (40%): Data analysis assignments to be handed in will be assigned three times during the semester (ILOs 1, 2, 3 , 4 and 5). 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. (ILOs 1.4, 1.5, 2.3, 2.4, 3.1, 3.2, 4.1, 5.1) Assignments (40%): Data analysis assignments to be handed in. (ILOs 1.4, 1.5, 2.3, 2.4, 3.1, 3.2, 4.1, 5.1)
Criteri di valutazione	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). • 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.
Bibliografia obbligatoria	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. Lämmle, T. (2025), Natural Language Processing & Web Analytics, Kindle Edition.
Bibliografia facoltativa	
Altre informazioni	
Obiettivi di Sviluppo Sostenibile (SDGs)	Istruzione di qualità

Modulo del corso

Titolo della parte costituente del corso	M1 - Statistical methods for high-dimensional data
Codice insegnamento	27512A
Settore Scientifico-Disciplinare	STAT-01/A
Lingua	Inglese
Docenti	prof. Davide Ferrari, Davide.Ferrari2@unibz.it https://www.unibz.it/en/faculties/economics-management/academic-staff/person/39001
Assistente	
Semestre	Primo semestre
CFU	6
Docente responsabile	
Ore didattica frontale	- 24 hours of in-person lectures

	- 12 hours of video lectures (counted as 24 hours to account for re-watching)
Ore di laboratorio	-
Ore di studio individuale	-
Ore di ricevimento previste	18
Sintesi contenuti	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.
Argomenti dell'insegnamento	• 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
Modalità di insegnamento	Recorded lectures, in-person teaching, exercises. 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 module 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.

Bibliografia obbligatoria	Lederer, J. (2022). Fundamentals of high-dimensional statistics. Springer International Publishing.
Bibliografia facoltativa	

Modulo del corso

Titolo della parte costituente del corso	M2 - Natural language processing and web analytics
Codice insegnamento	27512B
Settore Scientifico-Disciplinare	INFO-01/A
Lingua	Inglese
Docenti	prof. dr. Paul Michael Pronobis, Paul.Pronobis@unibz.it https://www.unibz.it/en/faculties/economics-management/academic-staff/person/49449
Assistente	
Semestre	Secondo semestre
CFU	6
Docente responsabile	
Ore didattica frontale	- 24 hours of in-person lectures - 12 hours of video lectures (counted as 24 hours to account for re-watching)
Ore di laboratorio	-
Ore di studio individuale	-
Ore di ricevimento previste	18
Sintesi contenuti	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.
Argomenti dell'insegnamento	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.
Modalità di insegnamento	Recorded lectures, in-person teaching, exercises. 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 module 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.
Bibliografia obbligatoria	Tunstall, L., Von Werra, L., & Wolf, T. (2022). *Natural language processing with transformers*. " O'Reilly Media, Inc. Lämmle, T. (2025), Natural Language Processing & Web Analytics, Kindle Edition.
Bibliografia facoltativa	