

# Syllabus

## *Descrizione corso*

<b>Titolo insegnamento</b>	Big data methods for economics and business
<b>Codice insegnamento</b>	27512
<b>Titolo aggiuntivo</b>	
<b>Settore Scientifico-Disciplinare</b>	NN
<b>Lingua</b>	Inglese
<b>Corso di Studio</b>	Corso di laurea magistrale in Data Analytics for Economics and Management
<b>Altri Corsi di Studio (mutuati)</b>	
<b>Docenti</b>	prof. Davide Ferrari, Davide.Ferrari2@unibz.it <a href="https://www.unibz.it/en/faculties/economics-management/academic-staff/person/39001">https://www.unibz.it/en/faculties/economics-management/academic-staff/person/39001</a> prof. dr. Paul Michael Pronobis, Paul.Pronobis@unibz.it <a href="https://www.unibz.it/en/faculties/economics-management/academic-staff/person/49449">https://www.unibz.it/en/faculties/economics-management/academic-staff/person/49449</a>
<b>Assistente</b>	
<b>Semestre</b>	Tutti i semestri
<b>Anno/i di corso</b>	2
<b>CFU</b>	12
<b>Ore didattica frontale</b>	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)
<b>Ore di laboratorio</b>	-

Ore di studio individuale	-
Ore di ricevimento previste	M1: 18 hours M2: 18 hours
Sintesi contenuti	<p>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.</p> <p>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.</p>
Argomenti dell'insegnamento	M1: <ul style="list-style-type: none"> <li>• High-dimensional data, big data and the curse of dimensionality</li> <li>• Convex criteria for model selection</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> M2:

	<ol style="list-style-type: none"> <li>1. Introduction to Natural Language Processing (NLP): Exploring the fundamentals of NLP, including its history, applications, and difference to other neural networks.</li> <li>2. Algorithmic Text Classification and Sentiment Analysis: Detailed instruction on various algorithms for categorizing text and extracting sentiment, comparing their effectiveness and use cases.</li> <li>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.</li> <li>4. Advanced Techniques in Information Retrieval: Utilization of cutting-edge neural network strategies combined with vector space models to efficiently retrieve information.</li> <li>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.</li> <li>6. Prompt Engineering for Enhanced Language Understanding: Crafting effective prompts to improve relation extraction, answer questions accurately, support dialog systems, and create responsive chatbots.</li> <li>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.</li> <li>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.</li> </ol>
<b>Parole chiave</b>	
<b>Prerequisiti</b>	
<b>Insegnamenti propedeutici</b>	
<b>Modalità di insegnamento</b>	<p>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</p>

	learning. The course is fully aligned with the principles of the Italian Universities Digital Hub (EDUNEXT) initiative ( <a href="https://edunext.eu">https://edunext.eu</a> ), which promotes the integration of digital resources and active learning strategies within university teaching.
<b>Obbligo di frequenza</b>	Recommended, but not required.
<b>Obiettivi formativi specifici e risultati di apprendimento attesi</b>	
<b>Obiettivi formativi specifici e risultati di apprendimento attesi (ulteriori info.)</b>	
<b>Modalità di esame</b>	<p>The overall exam mark will be determined by the assessment of the two modules (M1+M2).</p> <p>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.</p> <p>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.</p>
<b>Criteri di valutazione</b>	<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</li> </ul>

	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.
<b>Bibliografia obbligatoria</b>	<p>M1:</p> <p>Lederer, J. (2022). Fundamentals of high-dimensional statistics. Springer International Publishing.</p> <p>M2:</p> <p>Tunstall, L., Von Werra, L., &amp; Wolf, T. (2022). <i>Natural language processing with transformers</i>. " O'Reilly Media, Inc.</p>
<b>Bibliografia facoltativa</b>	
<b>Altre informazioni</b>	
<b>Obiettivi di Sviluppo Sostenibile (SDGs)</b>	

## *Modulo del corso*

<b>Titolo della parte costituente del corso</b>	M1 - Statistical methods for high-dimensional data
<b>Codice insegnamento</b>	27512A
<b>Settore Scientifico-Disciplinare</b>	SECS-S/01
<b>Lingua</b>	Inglese
<b>Docenti</b>	<p>prof. Davide Ferrari,            Davide.Ferrari2@unibz.it  <a href="https://www.unibz.it/en/faculties/economics-management/academic-staff/person/39001">https://www.unibz.it/en/faculties/economics-management/academic-staff/person/39001</a></p>
<b>Assistente</b>	
<b>Semestre</b>	Primo semestre
<b>CFU</b>	6
<b>Docente responsabile</b>	
<b>Ore didattiche frontali</b>	<ul style="list-style-type: none"> <li>- 24 hours of in-person lectures</li> <li>- 12 hours of video lectures (counted as 24 hours to account for</li> </ul>

	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	<ul style="list-style-type: none"> <li>• High-dimensional data, big data and the curse of dimensionality</li> <li>• Convex criteria for model selection</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>
Modalità di insegnamento	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 ( <a href="https://edunext.eu">https://edunext.eu</a> ), 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*

<b>Titolo della parte costituente del corso</b>	M2 - Natural language processing and web analytics
<b>Codice insegnamento</b>	27512B
<b>Settore Scientifico-Disciplinare</b>	INF/01
<b>Lingua</b>	Inglese
<b>Docenti</b>	prof. dr. Paul Michael Pronobis, Paul.Pronobis@unibz.it <a href="https://www.unibz.it/en/faculties/economics-management/academic-staff/person/49449">https://www.unibz.it/en/faculties/economics-management/academic-staff/person/49449</a>
<b>Assistente</b>	
<b>Semestre</b>	Secondo semestre
<b>CFU</b>	6
<b>Docente responsabile</b>	
<b>Ore didattica frontale</b>	<ul style="list-style-type: none"> <li>- 24 hours of in-person lectures</li> <li>- 12 hours of video lectures (counted as 24 hours to account for re-watching)</li> </ul>
<b>Ore di laboratorio</b>	-
<b>Ore di studio individuale</b>	-
<b>Ore di ricevimento previste</b>	18
<b>Sintesi contenuti</b>	<p>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,</p>

	equipping students to critically evaluate and implement state-of-the-art NLP solutions.
<b>Argomenti dell'insegnamento</b>	<ol style="list-style-type: none"> <li>1. Introduction to Natural Language Processing (NLP): Exploring the fundamentals of NLP, including its history, applications, and difference to other neural networks.</li> <li>2. Algorithmic Text Classification and Sentiment Analysis: Detailed instruction on various algorithms for categorizing text and extracting sentiment, comparing their effectiveness and use cases.</li> <li>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.</li> <li>4. Advanced Techniques in Information Retrieval: Utilization of cutting-edge neural network strategies combined with vector space models to efficiently retrieve information.</li> <li>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.</li> <li>6. Prompt Engineering for Enhanced Language Understanding: Crafting effective prompts to improve relation extraction, answer questions accurately, support dialog systems, and create responsive chatbots.</li> <li>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.</li> <li>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.</li> </ol>
<b>Modalità di insegnamento</b>	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 ( <a href="https://edunext.eu">https://edunext.eu</a> ), which promotes the integration of digital resources and active learning strategies within university teaching.
<b>Bibliografia obbligatoria</b>	Tunstall, L., Von Werra, L., & Wolf, T. (2022). <i>Natural language processing with transformers</i> . " O'Reilly Media, Inc.
<b>Bibliografia facoltativa</b>	