

CETIS - PhD TRANSVERSAL TRAINING ACTIVITY

Basic information of the PhD Training Activity	
Title	Fundamentals of Machine Learning
Professor	Dr. Cristina Montañaola Sales
Character	Transversal
Type of activity	Compulsory
Hours of class time	10 (5 sessions of 2 hours each)
PhD student level	1 st year students <input checked="" type="checkbox"/> 2 nd year students <input type="checkbox"/> 3 rd year students <input type="checkbox"/>
Teaching	Face-to-face <input type="checkbox"/> Face-to-face and online <input checked="" type="checkbox"/>
Semester	2
Goals	<p>This course is designed to provide a comprehensive introduction to the core concepts and techniques in Machine Learning (ML), tailored for PhD students with basic statistical knowledge but no programming experience. The primary aim is to equip students with a robust understanding of both Artificial Intelligence (AI) and ML, exploring their historical development, fundamental types, and diverse applications across various domains. Emphasis will be placed on using the Orange tool, a powerful yet user-friendly platform that allows for sophisticated data analysis and ML model implementation through visual programming, negating the need for coding skills. The course will guide students through the entire ML pipeline, from data acquisition and preprocessing to the selection and application of appropriate algorithms for both regression and classification tasks. Students will also learn how to evaluate model performance using various metrics and delve into clustering techniques to uncover patterns within data. Additionally, the curriculum includes advanced topics such as dimensionality reduction to manage high-dimensional datasets and text analysis for extracting insights from textual data. Ultimately, this course aims to build practical, hands-on skills, enabling students to apply ML methodologies effectively in their research and professional endeavours using the Orange platform.</p>
Contents	<ul style="list-style-type: none"> • Session 1: Introduction to AI & ML and Orange Platform Concepts Covered: <ul style="list-style-type: none"> – Overview of AI and ML: definitions, types, and applications. – Historical milestones in AI and ML development. – Introduction to the Orange platform: installation, interface, and basic functionalities. – Hands-on activity: Setting up Orange and exploring its basic features. • Session 2: Data Preprocessing and Initial Models Concepts Covered: <ul style="list-style-type: none"> – Data acquisition: sources and types of data. – Data preprocessing techniques: cleaning, normalization, and transformation. – Introduction to ML problem typologies: supervised vs. unsupervised learning. – Selecting and implementing basic ML models: regression and classification. – Hands-on activity: Load a dataset, preprocess it, and apply a simple classification model in Orange • Session 3: Performance Analysis and Clustering Concepts Covered:

	<ul style="list-style-type: none"> - Evaluation metrics for ML models: accuracy, precision, recall, F1-score, and ROC curves. - Introduction to clustering: k-means, hierarchical clustering, and their applications. - Practical usage of Orange for model evaluation and clustering analysis. - Hands-on activity: Evaluate the performance of a classification model and perform clustering analysis on a dataset using Orange. <ul style="list-style-type: none"> • Session 4: Dimensionality Reduction Concepts Covered: <ul style="list-style-type: none"> - The need for dimensionality reduction: curse of dimensionality. - Techniques: Principal Component Analysis (PCA), t-SNE, MDS and their implementation in Orange. - Hands-on activity: Apply PCA, MDS and t-SNE on a dataset and visualize the results in Orange. • Session 5: Text Analysis Concepts Covered: <ul style="list-style-type: none"> - Basics of text analysis and natural language processing (NLP). - Implementing text analysis in Orange: word clouds, sentiment analysis, and text classification. - Hands-on activity: Perform text preprocessing and analysis on a textual dataset using Orange.
<p>Learning outcomes</p>	<p>By the end of this course, students will be able to:</p> <ol style="list-style-type: none"> 1. Comprehend the Fundamentals of AI and ML: Grasp the basic concepts, types, and applications of AI and ML, and appreciate their impact on various sectors. 2. Navigate and Utilize the Orange Platform: Efficiently use Orange for data mining and ML tasks, leveraging its visual programming capabilities without requiring coding knowledge. 3. Preprocess and Analyse Data: Perform data cleaning, normalization, and transformation, and understand the importance of data preparation in ML workflows. 4. Apply and Evaluate ML Models: Implement and evaluate regression and classification models, using appropriate metrics to assess their performance. 5. Conduct Clustering and Dimensionality Reduction: Apply clustering techniques and dimensionality reduction methods to uncover patterns and reduce data complexity. 6. Perform Basic Text Analysis: Conduct text preprocessing and analysis, using Orange to extract insights from textual data.

Language

English

Assessment

- Participation and Attendance: Active involvement in sessions and consistent attendance (mandatory for a pass).
- Hands-on Activities: Completion of in-class practical exercises using Orange.
- Final Project: A mini-project involving the application of learned ML techniques on a dataset, demonstrating the ability to preprocess data, apply models, evaluate performance, and interpret results.