

# **SAIL - Split AI Lab School of Machine Learning and Data Science**

**Monday 26 May 2025 - Friday 6 June 2025**

**UNIST Technological Park, Croatia**

## **Scientific Programme**

The Split AI Lab is an intensive two-week program designed to provide participants with a strong foundation in Machine Learning and Data Science, combining theoretical instruction with hands-on implementation. The curriculum is structured to gradually build knowledge, covering fundamental statistical concepts, essential machine learning algorithms, and modern deep learning techniques.

The program is divided into three main components: supervised learning, unsupervised learning, and deep learning, with each section emphasizing both theoretical understanding and practical application.

### **Core Topics Covered**

#### *Statistical Foundations & Data Preprocessing*

essential statistical concepts and learning best practices in feature engineering, dimensionality reduction, and data preprocessing techniques.

#### *Supervised Learning*

classification and regression models, including k-Nearest Neighbors (kNN), Decision Trees, Linear and Logistic Regression, Support Vector Machines (SVMs), and Ensemble Methods. Emphasis is placed on model evaluation, cross-validation, and trade-offs between different performance metrics.

#### *Unsupervised Learning & Clustering Techniques*

k-Means Clustering, Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Singular Value Decomposition (SVD), along with methods for analyzing high-dimensional data.

#### *Advanced Topics & Model Optimization*

kernel methods, regularization techniques, non-linear regression models, and an introduction to structured machine learning models.

#### *Deep Learning & Neural Networks*

fundamental neural network architectures, including feedforward networks, convolutional neural networks (CNNs), and basic deep learning frameworks.

### **Hands-on Learning Approach**

Throughout the program, participants will implement these models using Jupyter Notebooks and standard Python libraries such as NumPy, Pandas, Scikit-learn, and TensorFlow/PyTorch. Practical applications will include real-world datasets and Kaggle-style challenges, allowing participants to gain experience in model training, tuning, and performance evaluation.