

Machine Learning for SW Developersd - 70 academic hours

Workshop Overview:

The Machine Learning workshop provides hands-on training in supervised and unsupervised learning using Python and Scikit-learn. Participants build end-to-end ML pipelines covering feature engineering, model selection, and evaluation, while working with core algorithms such as regression, decision trees, random forests, k-NN, and clustering.

The workshop is based on real-world business use cases and focuses on practical techniques for addressing common ML challenges, including overfitting, imbalanced data, and anomaly detection, with a strong emphasis on real-world application.

What Participants Gain?

- Hands-on ability to build end-to-end Machine Learning pipelines using Python and Scikit-learn
- Practical experience selecting features, training models, and evaluating performance
- Applying core ML algorithms to real-world business use cases
- Handling common ML challenges such as overfitting, imbalanced data, and anomaly detection
- Designing production-ready ML solutions with reproducible workflows

Target Audience

- Data Scientists and ML Engineers working with cloud-based ML systems
- Data Engineers responsible for cloud data pipelines and ML infrastructure
- DevOps and Platform Engineers supporting AI and Machine Learning workloads
- Developers building cloud-based ML applications
- Data and technology professionals seeking to adopt cloud-native MLOps practices

Prerequisites

- Basic familiarity with Machine Learning concepts and workflows
- Experience with at least one programming language, preferably Python
- General understanding of data processing and ML pipelines
- Basic familiarity with cloud or DevOps concepts is an advantage
- Good technical English proficiency

Workshop Syllabus

Concepts

- Supervised & unsupervised learning
- Pipelines – Transformers & Estimators
- Feature engineering
- Dimensionality reduction
- Model selection – Cross-validation & grid search
- Overfitting & regularization
- Ensemble methods – Voting, bagging & boosting
- Imbalanced data
- Anomaly detection
- Clustering
- Metrics and similarities
- Scoring

Models

- Linear regression
- Logistic regression
- Decision trees (inc. random forest)
- K-nearest neighbors (k-NN)
- K-means
- Agglomerative clustering