

AA-DM101 Industrial Data Mining



With the advent of the connected factory, in the Industry 4.0 paradigm, the **effective management and knowledge extraction** from field-level data is becoming a critical challenge. The course focuses on a systematic approach and techniques for **managing large data sets, building and improving predictive models** for solving practical issues. This leads to defining and answering key engineering and management questions by having control over the tuning aspects and design decisions behind high level machine learning tools and frameworks. Core topics include: data mining methodology, data handling and structuring, modeling techniques for regression and classification, model selection, validation and

operationalization. Hands-on sessions and implementation of data mining pipelines, using specific tools e.g. RapidMiner, bring considerable productivity enhancements. These include **relevant practical examples** for predictive maintenance of equipment and enhancement of production line performance in industrial manufacturing environments.

Target audience:

Data scientists, Maintenance and Quality Engineers, IT Staff

Prerequisites:

Technical or quantitative background, basic skills for handling structured data e.g. Excel

Contents:

- CRISP-DM data mining methodology
- Preparation of heterogenous data sets for data mining algorithms
- Classification of data mining tasks: supervised and unsupervised learning, regression and classification methods
- Evaluation of model accuracy, the bias-variance trade-off
- Hands-on sessions – Introduction to RapidMiner
 - Data ETL
 - Linear regression techniques
 - Classification by means of k-NN algorithm
 - Decision trees
- Practical example: predictive maintenance based on equipment failure models
- Practical example: production line performance improvement

Objectives:

- During this course the attendees will acquire knowledge and practical abilities related to:
- Structured approach to handling large data sets from various sources (raw data, machine logs, aggregated reports etc.)
 - Solve specific maintenance and production problems and improve performance by applying data science processes
 - Correctly select the appropriate modeling technique based on available data and domain knowledge
 - Evaluation and implementation of the developed models in the daily workflow
 - Essential visualization, reporting and presentation skills of the results of the data analysis
 - Basic introduction to RapidMiner Studio for advanced analytics and data science workflows

Duration: 3 days

Course breakdown *draft*

Day 1

- Module 1: Introduction; Data mining methodology; Application and relevance in the current industrial context; Types of problems that can be solved using DM.
- Coffee break
- Module 2: Approaching large heterogeneous data sets; data types and preparation for mining tasks.
- Lunch
- Hands-on session 1: RapidMiner Studio environment, main components and process structure.
- Coffee break
- Hands-on session 2: Data ETL using RapidMiner Studio; Basic statistical analysis of input data.

Day 2

- Module 3: Classification of data mining tasks: supervised and unsupervised learning, regression and classification methods.
- Coffee break
- Module 4: Evaluation of model accuracy, the bias-variance trade-off.
- Lunch
- Hands-on session 3: Supervised learning; Linear regression techniques for prediction of continuous outputs; Classification using k-NN.
- Coffee break
- Hands-on session 4: Decision trees

Day 3

- Hands-on session 5: Predictive maintenance for averting machine failures*
- Coffee break
- Hands-on session 6: Production line performance improvement; Case study – mechanical parts*
- Lunch
- Hands-on session 7: Production line performance improvement; Case study – electronic modules*
- Coffee break
- Interactive session: Structure and scope of popular data science platforms and open-source projects; Best practices for interdisciplinary data science teams; QA; Next steps.

The course takes place in the Asti Automation Training Center in Bucharest. All required materials are provided: laptop with pre-installed software and data sets, printed hand-outs, slides, etc. Coffee breaks and lunches are included. On-premise sessions can be arranged for minimum groups of 6 participants.