https://www.projectmates.eu/

MATES ED2MIT

Education and Training for Data Driven Maritime Industry

Data Science and Analytics Foundation (DSAF)

Introduction: MATES Project and course overview

Yuri Demchenko MATES Project

University of Amsterdam

Maritime Alliance for fostering the European Blue economy through a Marine Technology Skilling Strategy

Co-funded by the Erasmus+ Programme of the European Union

DSAF Course Overview

Conditions

- This course is provided as self-training/self-study course
  - No regular tutor support provided
- Look for announcements for tutor-lead training
- Course materials are available at EDISON Community website

Co-funded by the
Erasmus+ Programme
of the European Union

Course Topics - Tutorials

- Introduction to course and Project MATES
- Tutorial D01: Research Methods in Data Science
- Tutorial D02: Statistical Data Analysis Basics: Data Structures, Statistical Characteristics
- Tutorial D03: Data Preparation and Processing
- Tutorial D04: Data Analysis Principles and Techniques, Exploratory Data Analysis
- Tutorial D05: Qualitative Data Analysis
- Tutorial D06: Machine Learning: Classification Methods
- Tutorial D07: Machine Learning: Cluster Analysis Overview

Course Practice

- Practice 01: Data cleaning and preparation with Python Pandas
- Practice 02: Python for statistical data analysis and visualisation
- Practice 03: Data preparation and exploration with RapidMiner
- Practice 04: Data analysis and Machine Learning with RapidMiner

Goal

- Provide introductory overview of the foundation of the Data Science and Analytics
  - To be used as a reference in the future study or work
- Provide basic examples of using Python and RapidMiner for simple data analysis tasks
  - Provide a starting point for further developing practical skills with the data analysis tools

Expected commitment

- This course is built of the best examples of existing learning materials and courses
- Lectures (in a form of tutorials) provide a brief overview of the main foundational topics and are intended as a reference material for further deeper study
- Although this course provides basic examples of using Python and RapidMiner for data analysis, it may require efforts from participants to go deeper and use preliminary courses, in particularly for Python
- Working with examples will require the set-up of a comprehensive working environment and the search for additional guidelines on the Internet
- Working with examples may produce errors and requires debugging programs. Be ready to experiment and search the web for solutions

Yuri Demchenko,
Senior Researcher, Lecturer, Complex Cyber Infrastructure Group, UvA

- Graduated and PhD from National Technical University of Ukraine “Kiev Polytechnic Institute”
  - University of Amsterdam – since 2003
- Research areas
  - Big Data Infrastructure and Data Science platforms
  - Cloud architecture framework, DevOps and cloud automation platform
  - Cloud security and compliance
- Teaching courses (on campus and online)
  - **Big Data Infrastructure and Technologies for Data Analytics (BDIT4DA)**
  - Cloud Computing Engineering (CCENG), Security Engineering (SECENG))
  - DevOps and Cloud based Software Development (DevOps)
  - **Web technologies and Databases (WebDB)**
  - **Data Science and Analytics Foundations (DSAF), Professional Issues in Data Science**
- Recent projects
  - EDISON: Building the Data Science Profession for Europe
  - **MATES: Digitalisation of the European Blue Economy**
  - FAIRsFAIR/EOSC: FAIR Data Management and Data Stewardship
  - GEANT4 Research: Cloud aware networking infrastructure provisioning on-demand
  - SLICES-DS: Research Infrastructure for ICT

Digital and Data Training: General goals

- Meet maritime experts and community to obtain feedback on how to deliver digital and data technologies to the wide community of maritime experts and community – by achieving digital and data literacy
  - Goal: Obtain feedback from professional maritime community via interactive teaching
  - Potentially find cooperators to transfer knowledge: train the trainers
- Provide basic information on Big Data and Data Analytics technology and tools
- Introduce to Data Management and Governance practices
- Build trust in data-driven and digitalized operations by providing assurance of data quality, algorithms, sensors, systems and cyber security
- Facilitate digital readiness and transformation, understand benefits

MATES ED2MIT Course on Data Governance and Data Management

Second in series of ED2MIT training webinars

Course material can be found online

- Provide general information on Data Governance and Management in the context of the European Data Governance policy
- Introduce Data Management best practice and Data Management Plan template

Industry 4.0 and demand for new skills

The Fourth Industrial Revolution, which includes developments in previously disjointed fields such as

- artificial intelligence and machine-learning, robotics, nanotechnology, 3-D printing, and genetics and biotechnology,
- will cause widespread disruption not only to business models but also to labour markets over the next five years,
- with enormous change predicted in the skill sets needed to thrive in the new landscape.
- Ref report *The Future of Jobs*, published by the World Economic Forum.

Maritime Industry Digital Transformation and Skills Strategy – Toward Industry 4.0

Digital Transformation

- Digitalisation and IoT
- Intelligent Information
- Data Management
- Digital Assets Manage
- Data Driven Optimisation
- Agile Continuous Improvement
- Customer Experience
- People and skills

Digital Competences/Skills

- Automation, robotics, electrical vehicles
- Information and data literacy
- Communication and collaboration
- Digital content creation, safety
- Problem solving and critical thinking

Industry 5.0 – Human centric Industry

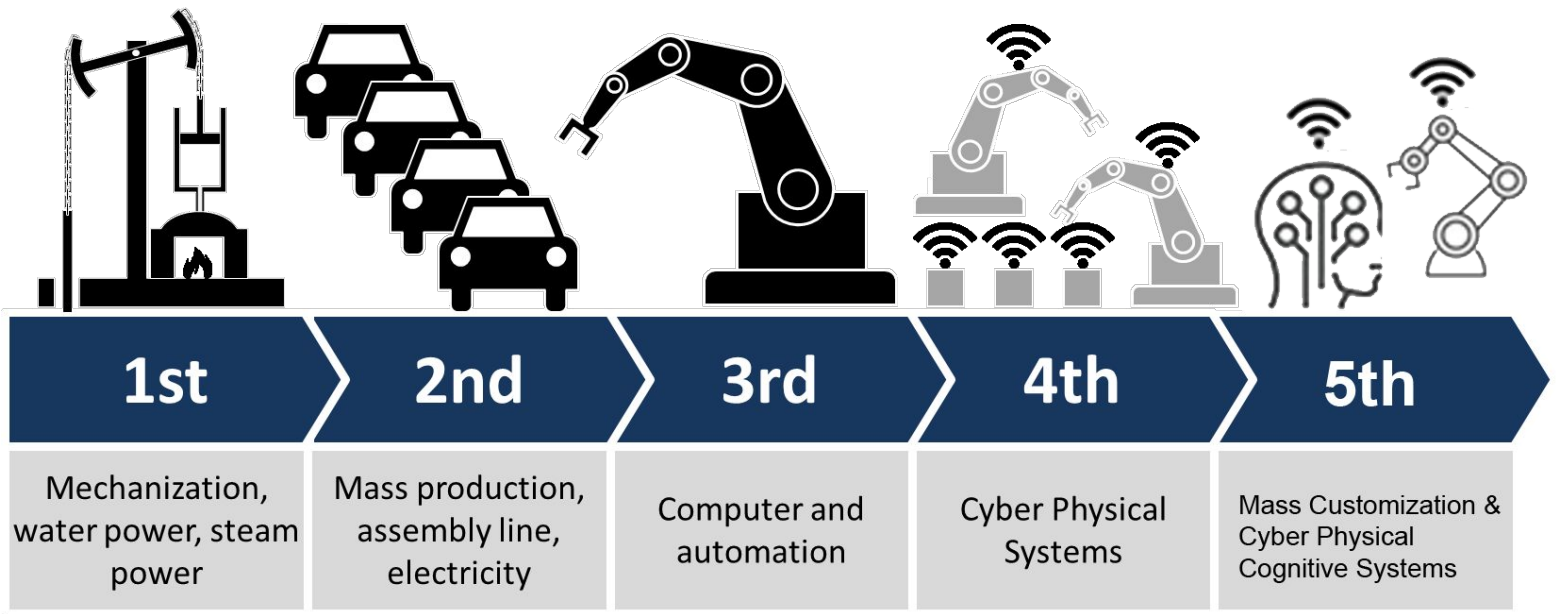
- The term Industry 5.0 refers to people working alongside robots and smart machines
  - Humans and robots co-working (cobots)
  - It's about robots helping humans work better and faster by leveraging advanced technologies like the Internet of Things (IoT) and big data
    - Robots will learn from humans, as well as sharing their ability to do the tasks operators can't, or don't need to do.
  - It adds a personal human touch to the Industry 4.0 pillars of automation and efficiency
  - Human inspired design: Mass personalization can be achieved when human touch returns to manufacturing
  - Industry 5.0 is the re-humanisation of the automation race

#1 Industry 5.0 is aimed at supporting – not superseding – humans

#2 Industry 5.0 is about finding the optimal balance of efficiency and productivity

#3 The progress of Industry 5.0 is unavoidable

Reference: Industry 5.0 | European Commission (europa.eu)



- Industry 5.0 is characterised by going beyond producing goods and services for profit. It shifts the focus from the shareholder value to stakeholder value and reinforces the role and the contribution of industry to society.
- It places the wellbeing of the worker at the centre of the production process and uses new technologies to provide prosperity beyond jobs and growth while respecting the production limits of the planet.
- It complements the existing "Industry 4.0" approach by specifically putting research and innovation at the service of the transition to a sustainable, human-centric and resilient European industry.

Industry 5.0: Towards more sustainable, resilient and human-centric industry

- Industry 5.0: Towards more sustainable, resilient and human-centric industry
- https://ec.europa.eu/info/news/industry-50-towards-more-sustainable-resilient-and-human-centric-industry-2021-jan-07_en

- Elements pertinent to Industry 5.0 are already part of major Commission policy initiatives
  - adopting a human-centric approach for digital technologies including artificial intelligence (AI White Paper)
  - **up-skilling and re-skilling European workers, particularly digital skills** (Skills Agenda and Digital Education Action plan)
  - modern, resource-efficient and sustainable industries and transition to a circular economy (Green Deal)
  - a globally competitive and world-leading industry, speeding up investment in research and innovation (Industrial Strategy)

Industry 5.0 website

- Industry 5.0 website https://ec.europa.eu/info/research-and-innovation/research-area/industrial-research-and-innovation/industry-50_en
- Industry 5.0 is characterised by going beyond producing goods and services for profit. It shifts the focus from the shareholder value to stakeholder value and reinforces the role and the contribution of industry to society.
- **Industry 5.0** – Report January 2021
  - https://op.europa.eu/en/publication-detail/-/publication/468a892a-5097-11eb-b59f-01aa75ed71a1/
  - Towards a sustainable, human-centric and resilient European industry
- **Enabling Technologies for Industry 5.0** - Sept 2020
  - https://ec.europa.eu/info/publications/enabling-technologies-industry-50_en
  - Results of a workshop with Europe's technology leaders

Related: Society 5.0 (according to Japan)
https://www8.cao.go.jp/cstp/english/society5_0/index.html

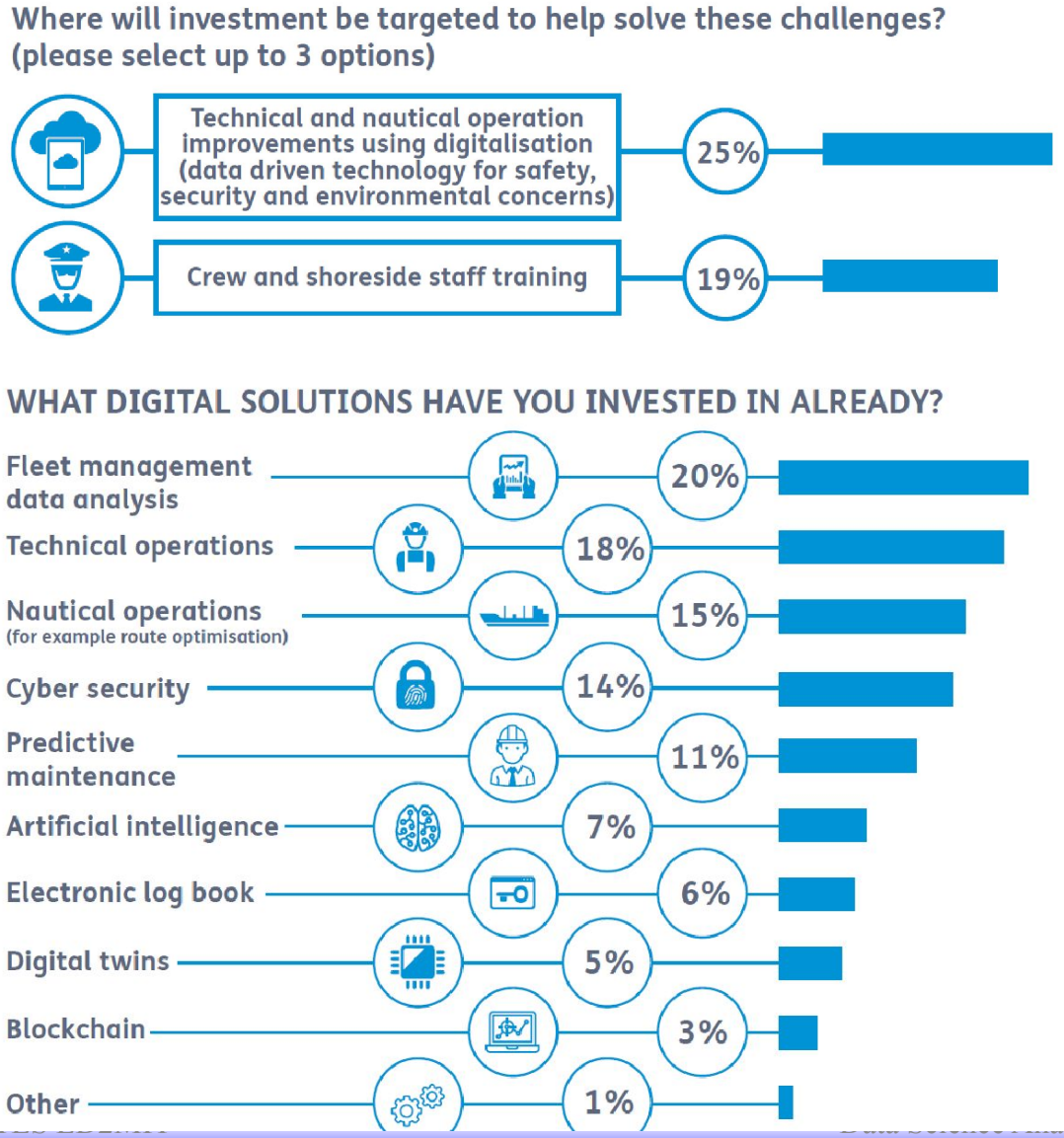
- Society 5.0 was proposed in **the 5th Science and Technology Basic Plan** as a future society that Japan should aspire to.
  - It follows the hunting society (Society 1.0), agricultural society (Society 2.0), industrial society (Society 3.0), and information society (Society 4.0).
- Definition: "A **human-centered society** that balances economic advancement with the resolution of social problems by a system that highly integrates cyberspace and physical space."

Stages of digitalization in maritime transport

The effects of digitalization on maritime transport can be divided into the following three stages:

1. Optimization – maximizing efficiency and reliability in existing processes to reduce the costs of trading
2. Extension – moving beyond efficiency to produce opportunities for new services and businesses
3. Transformation – reinventing logistics, trade and business models, based on data-driven revenue streams and shifts in trade flows

Digitalisation is game changer in shipping (Wartsila report)



- The benefit of technology is in value it brings to the business, and that value is in proportion to Leaders' understanding
- Staff education and training is a key to adopt new technologies

[Ref]https://lloydslist.maritimeintelligence.informa.com/LL1128397/Click-and-connect-to-explore-shippings-gamechanger *Source: Shinonome Production/Shutterstock.com*

Digitalisation in Shipping

- *“The analogy here is that we are still sending data in a break bulk mentality with myriad standards, formats and sizes. The birth of a single global standard — the 20 ft container — that revolutionised shipping has not yet been found digitally.”*
  https://lloydslist.maritimeintelligence.informa.com/LL1136038/Digitalising-shipping-Whats-the-hold-up
- Standardisation is a key to successful digitalisation

Webinar: How to digitalise shipping, 17th March 2021

Digital Transformation in Renewable Energy

https://nexusintegra.io/digital-transformation-renewable-energy/

- Digitalisation tools and platforms help **build renewable energy plants** with automated processes, for informed decision making
  - Basis for a **more decentralised generation**, thus avoiding isolated 'energy islands'
- These platforms **reduce downtime** by offering alerts based on predictive maintenance, anticipating asset maintenance
- Digital platforms allow a **more accurate forecast of the weather and market conditions**, which helps to maximise renewable production, by offering a deep analysis of all information received in real time, enabling decision-making and offering stability on demand.
- The use of artificial intelligence and machine learning to **optimise the engineering and construction of new renewable sources and plants** reduces time to market, anticipating the benefits of free C02 generation and increasing production.
- Enabling technologies
  - Big Data
  - Digital Twins
  - Edge networks and clouds

Opportunities with Digitalisation

- Cross-sector sharing
- Catch up with the IoT data tsunami
  - Sensors network
  - Robotics
  - Vehicles connectivity
- Starting data pipeline from IoT and edge computing
  - Early problems identification

Data driven culture in enterprise

Building a Data-Driven Culture in Enterprise: Introducing the 4 Pillars

- **Data maturity.** Solid data maturity is foundational to a data culture.
  - Organization's data maturity manifests itself in every individual in your organization having an easy and appropriate level of access to the prepared and accurate data they need
  - Importance of a well-defined CDO role and other related roles
- **Data-driven leadership.** Leaders define the culture of their organization.
  - A data-driven leader supports a culture of data by demonstrating data-driven decision making and involve the team members
  - A data-driven leader sees data as a strategic asset and makes "think and act data" a key strategic priority
- **Data literacy.** Individual decision makers must be data literate to leverage their data
  - The CDO office needs to invest in enterprise wide data literacy, where every role is upgraded with the right level of data science skills
- **Data-driven decision-making processes.** Establish a structured process of forward-looking decision making and backward-looking reviews of decisions.
  - Build experience of aligning data analytics, insight and data-driven decision-making processes

Five Characteristics of a Data-Driven Company

#1. Creative executives who run their businesses with passion and curiosity

- Being data-driven requires a bit of a researcher's mindset – the curiosity to dig into the data and glean insights from it that can be of use for the business.

#2. Data democratization

- Data-driven organizations emphasize the importance of broad data access for all employees.

#3. Data literacy

- An organization's ability to succeed in the digital era is heavily dependent on its employees' data literacy: the ability to read, work, analyze, and argue with data.
- Example of how to respond to a data literacy problem is Data University at Airbnb. Airbnb could not have a data scientist in every room to inform every decision with data.

#4. Automation of data management workloads

- A core criterion for a data-driven organization is how much data analytics tools are automated and provide information in a form that can be easy for decision making

#5. A companywide, data-driven culture

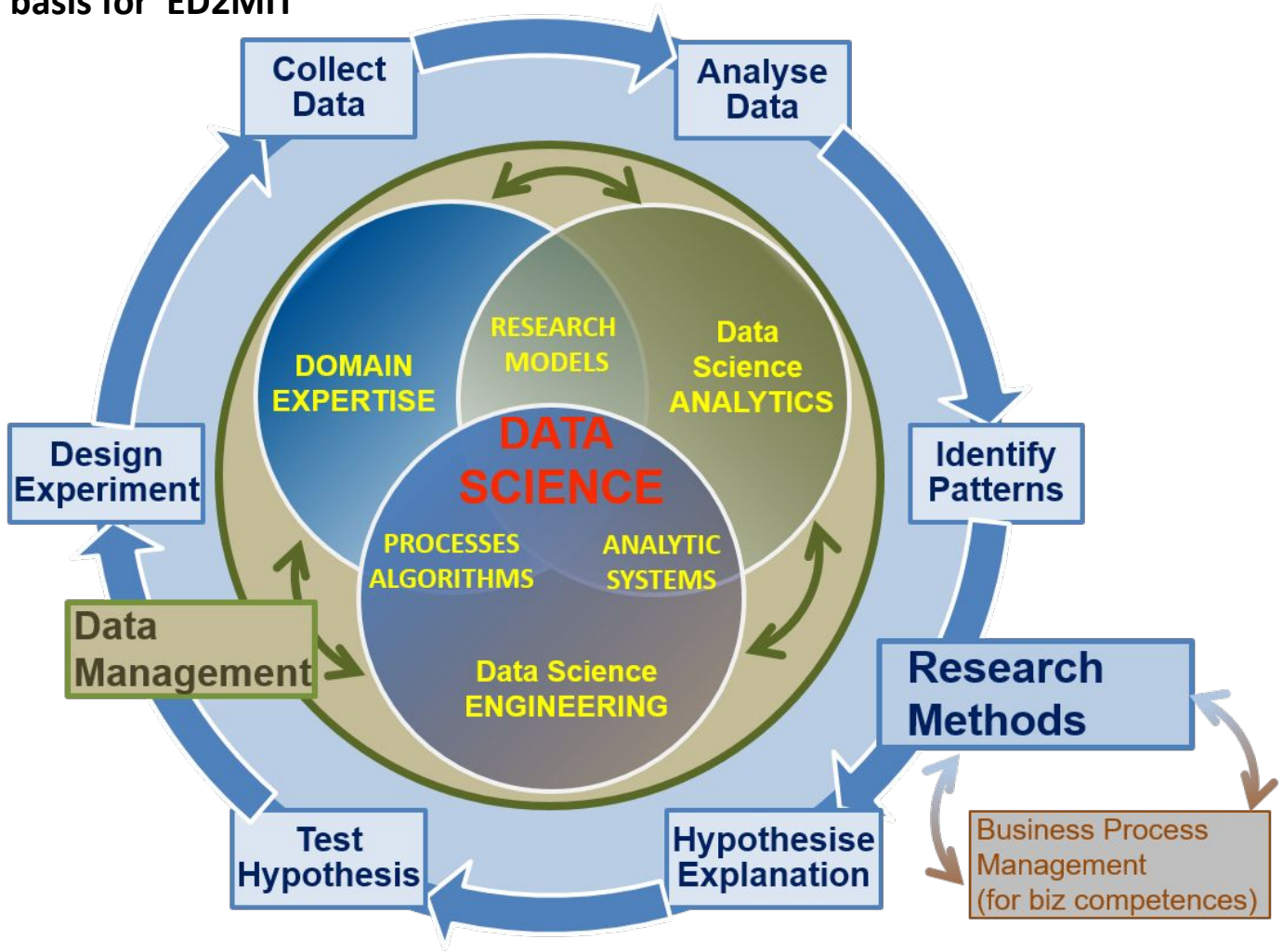
- Becoming data-driven involves more than technology and tools. It also requires a shift in the enterprise's mindset and culture.
- IBM example: Require all analytics blending with open data (social media, weather, climate etc.)

MATES Project contribution

- Study and report on demanded skills in maritime industry
- Pilot experiences and Roadmap
  - Pilot Experience ED2MIT - Education and Training for Data Driven Maritime Industry
- Leverage EDISON Data Science Framework (EDSF) that created a basis for Big Data, Data Science and digital and data skills education and training
  - EDSF provides a basis for ED2MIT training program
- Leverage European Digital Competence framework DigCom2.1
- Provide training on digital and data technologies
  - Train trainers, create pool of reference training materials

EDISON Project (2015-2017): Data Science Competence Groups - Research

EDISON Data Science Framework (EDSF) provides a basis for ED2MIT



Data Science Competences include 5 groups

- Data Science Analytics
- Data Science Engineering
- Domain Knowledge and Expertise
- Data Management
- Research Methods and Project Management
  - Business Process Management (biz)

Scientific Methods

- Design Experiment
- Collect Data
- Analyse Data
- Identify Patterns
- Hypothesis Explanation
- Test Hypothesis

Business Operations

- Operations Strategy
- Plan
- Design & Deploy
- Monitor & Control
- Improve & Re-design

Acknowledgement

- This work is supported by the ERASMUS+ MATES project
- The work is committed to the Open Source under Creative Commons 4.0 CC BY License

Additional materials

- ED2MIT Training on digital and data skills (accessible online)

ED2MIT: Digital and Data Competence Groups

A. Data – B. Cloud – C. Digital Content – D. Data Science & Analytics

A. Data related competences and technologies
B. Cloud services and cloud economics
C. Digital content creation, access and management
D. Data Science and Big Data Analytics

A. Data related competences and technologies

A. Data – B. Cloud – C. Digital Content – D. Data Science & Analytics

A. Data related competences and technologies

A.1. Big Data definition and technologies: 6V of Big Data and challenges for companies. Big Data examples from research and industry
A.2. Data collection, access and sharing
A.3. Data formats, data models, metadata
A.4. Data Storage and databases, SQL scripting and simple commands
A.5. Data inspection, Data protection, data backup and archiving
A.6. Cloud based services and tools for data storage, sharing and management
A.7. Open Data repositories, test datasets, developer communities
A.8. Organisational and private Data Management, FAIR Data Principles, organisational roles, Data Stewards

B. Cloud services and cloud economics

A. Data – B. Cloud – C. Digital Content – D. Data Science & Analytics

B. Cloud services and cloud economics

B.1. Cloud service models: IaaS, PaaS, SaaS, Apps. Use examples and cloud Service Providers. Cost model of cloud services.
B.2. Company IT infrastructure migrating to cloud: benefits and problems
B.3. Cloud and Big Data, cloud based Big Data platform and services
B.4. Data storing, backing up, sharing and processing in clouds (for organisational and private data)
B.5. Practical exercises with cloud services: Cloud management console and cloud services deployment and access.

C. Digital content creation, access and management

A. Data – B. Cloud – C. Digital Content – D. Data Science & Analytics

C. Digital content creation, access and management

To be acquired as self-study or expected as known.

C.1. Internet and World Wide Web
C.2. HTML, CSS, JavaScript for active pages
C.3. UX design and web portal services
C.4. Web applications security
C.5. PHP and interactive web sites (advanced)

To be provided as self-study materials

D. Data Science and Big Data Analytics

A. Data – B. Cloud – C. Digital Content – D. Data Science & Analytics

D. Data Science and Big Data Analytics

This course is provided as a general overview of the listed below topics. More in-depth training and learning will require more time commitment and pre-requisite knowledge.

D.1. Statistical methods and Probability theory
D.2. Data description and Statistical Data Analysis
D.3. Data preparation: data loading, data cleaning, data pre-processing, parsing, transforming, merging, and storing data
D.4. Qualitative and Quantitative data analysis
D.5. Classification: methods and algorithms
D.6. Cluster analysis basics and algorithms
D.7. Performance of data analytics algorithms and tools
D.8. Visualizations of data analysis and dashboards
D.9. Organizing data analytics process following CRISP-DM and Data Science Process