DS&AI

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

Curriculum Development in Data Science and Artificial Intelligence

599600-EPP-1-2018-1-TH-EPPKA2-CBHE-JP

Deliverable 2.5: DS & AI Course Outlines

Master and Professional Courses

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

PROJECT INFORMATION

Acronym	DS&AI
Project Title	Curriculum Development in Data Science and Artificial Intelligence
Contract Number	599600
Start Date	15 Nov 2018
Duration	36 months

DELIVERABLE INFORMATION

Deliverable Number	[2.5]
Deliverable Title	DS & AI course outlines
Submission Due Date	March 2020
Actual Submission Date	July 2020
WP Number and Title	WP 2.5, *DS & AI Course Outlines*
WP Lead Beneficiary	[AUEB- AMC]
Author and Organization	[K. Fraidaki- M. Skiada] AUEB
Dissemination Type	Report
Dissemination Level	All Project Partners
Quality Reviewer 1	Chutiporn Anutariya, AIT
Quality Reviewer 2	Taufik Fuadi Abidin/UNSYIAH
First Quality Review Date	23 August 2020
Quality Review Pass Date	31 August 2020

DISCLAIMER

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Contents

Abstract

The MS in Data Science and AI (DS&AI) is an intensive full-time Master degree program aiming to give students a strong base in the computational and mathematical foundations of data science and artificial intelligence as well as to give ample opportunities to explore and specialize in various application areas. The curriculum consists of a common set of core courses, a collection of electives exploring advanced topics and applications, and a significant one-year research project leading to a master's thesis. The same logic is followed in the design of professional courses. The outlines of the of core, elective and professional courses have been co-developed by all the partners of the consortium and in this deliverable are presented.

Introduction

During the deliverable 2.5 the outlines of 19 courses (16 master and 3 professional courses) had to be designed according to the proposal. The selected courses have been decided by the consortium to be the following:

Master Courses

- Core Courses
  1. Artificial Intelligence
  2. Data Modeling & Management (DMM)
  3. Business Intelligence and Analytics
  4. Machine Learning
  5. Computer Programming for Data Science and Artificial Intelligence
- Elective Courses
  1. Computational Linguistics
  2. Computer Vision
  3. Distributed Systems
  4. Human Computer Interaction and Information Visualisation
  5. Knowledge Representation
  6. Multicriteria Optimization and Decision Analysis
  7. Nature-Inspired Computing
  8. Recent Trends in Machine Learning
  9. Software Development and Project Management for DS & AI
  10. Spatial-Temporal Data Analysis
  11. Social Network Analysis (SNA)

Professional Courses

1. Real World Big Data Engineering
2. NoSQL Database Design and Development
3. Machine Learning and Professionals

The outlines were co-created with the contribution of all partners. All the working groups were created having representatives of all universities. Each working group was responsible of one course and had sufficient time to discuss in detail and then decide the outlines. After all the working groups had developed the first version of each course outline, they presented their suggested outlines to the entire consortium to collect feedback and to avoid content overlays. Athens University of Economics and Business coordinated the process and formatted the outcomes.

Core Courses

ARTIFICIAL INTELLIGENCE

Credits: 3

Objective: Introducing students to fundamentals of Artificial Intelligence. Students will be exposed to several techniques on planning and decision procedures ranging from precise to uncertain and temporal reasoning with applications to intelligent agents.

Learning Outcomes: Students, on successful completion of the course, will be able to

1. Demonstrate fundamental insights into practical planning and decision procedures;
2. Reason under uncertainty;
3. Apply planning techniques into intelligent agents.

Prerequisites: None.

Course Outline:

I. Introduction to AI
    1. What is (artificial) intelligence
    2. History of artificial intelligence

II. Intelligent agents

III. Planning and decision
    1. Decision trees and searching techniques
    2. Heuristic algorithms

IV. Constrained planning
    1. Recall on propositional and predicate logic
    2. Unification and resolution
    3. Prolog and/or constraint solvers

V. Planning under uncertainty
    1. Bayesian networks
    2. (Partially observable) Markov decision networks

VI. Temporal planning
    1. Temporal reasoning
    2. Scheduling

Laboratory Session(s): 30 hours of lab sessions for assignments on topics III, IV, V and VI and project using a logic or a constraint programming language such as Prolog.

Learning Resources:

Reference Books:

- Russel, S., & Norvig, P. (2013). *Artificial intelligence: a modern approach*. Pearson Education Limited.
- Ghallab, M., Nau, D., & Traverso, P. (2004). *Automated Planning: theory and practice*. Elsevier.
  Bratko, I. (2001). *Prolog programming for artificial intelligence*. Pearson education.

Teaching and Learning Methods:

1. *Direct instruction* based teaching for the lectures using visual aid via slides.
2. *Self-learning* method for Prolog based on a small introduction by the teacher.
3. *Inquire based* teaching for the labs through one medium scale project and assignments related to section III to VI in the course outline. Assignments should be done using Prolog.

Evaluation Components

1. Exam
2. Assignments
3. Project

DATA MODELING & MANAGMENT (DMM)

Credits: 7.5 ECTS

Objective: The course emphasizes on emerging data models and technologies suitable for managing different types and characteristics of data. Student will develop skills for analyzing, evaluating, modeling and developing database applications with concerns on both technical and business requirements.

Learning Outcomes: Students, on successful completion of the course, will be able to

1. Explain data modeling and management concepts.
2. Design and organize various types of data using a relational and non-relational data models.
3. Analyze the characteristics and requirements of data and select an appropriate data model.
4. Identify, implement and perform frequent data operations (CRUD: create, read, update and delete) on relational and NoSQL databases.
5. Describe the concepts and the importance of big data, data security, privacy and governance.
6. Describe the concepts and the importance of data engineering and data visualization.

Prerequisites: None

Course Outline:

I. Recall: Relational Data Model and Management
   1. Relational Model Concepts
   2. SQL
   3. Relational Database Design and Normalization
   4. Relational Database Management Systems (RDBMSs)

II. NoSQL Data Modeling and Management
   1. NoSQL Concepts and Characteristics
   2. Major Categories of NoSQL Data Models
   3. NoSQL Database Design

4. NoSQL Features and Operations

III. Data Distribution
1. Data Sharding and Replication Models
2. CAP Theorem

IV. Transaction Processing and Consistency Models
1. Transaction Processing Concepts
2. ACID Model
3. BASE Model

V. Large Scale Data Handling
1. Big Data characteristics
2. Big Data Modeling and Management

VI. Applications and Case Studies

VII. Data Engineering
1. Business Understanding
2. Data Acquisition and Understanding
3. Data Cleansing
4. Data Preparation, Transformation and Feature Engineering

VIII. Introduction to Related Topics
1. Data Security
2. Data Privacy and Legal Issues,
3. Data Governance: Social and Ethical Issues, Biasness (gender, religions, etc.)

Laboratory Session(s): 30 hours of laboratory sessions of NoSQL data stores, tools, CRUD operations, and API development/usage.

Learning Resources:

Textbooks:

Meier, A., & Kaufmann, M. (2019). *SQL & NoSQL Databases*. Springer Fachmedien Wiesbaden.

Reference Books:

- Kleppmann, M. (2017). Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems. " O'Reilly Media, Inc.".
- Sullivan, D. (2015). NoSQL for mere mortals. Addison-Wesley Professional.

- Sadalage, P. J., & Fowler, M. (2013). NoSQL distilled: a brief guide to the emerging world of polyglot persistence. Pearson Education.
- Perkins, L., Redmond, E., & Wilson, J. (2018). Seven databases in seven weeks: a guide to modern databases and the NoSQL movement. Pragmatic Bookshelf.
- Guy, H. (2015). Next generation databases: NoSQL, newSQL, and big data.
- Robinson, I., Webber, J., & Eifrem, E. (2015). Graph databases: new opportunities for connected data. " O'Reilly Media, Inc.".
- Navathe, S. B., & Elmasri, R. A. (2001). Fundamentals of Database Systems with Cdrom and Book. Addison-Wesley Longman Publishing Co., Inc..

Journals and Magazines:

- IEEE Transactions on Knowledge and Data Engineering
- ACM Trans. Database Systems.
- ACM Trans. Information Systems.

Teaching and Learning Methods:

1. Lecture and Lab
2. Assignment
3. Project
4. Discussions and Case Studies

Evaluation Components

1. Examination
2. Group Work
3. Individual Work
4. Others

BUSINESS INTELLIGENCE AND ANALYTICS

Credits: 3

Objective: Business intelligence (BI) is a process of analyzing business data to obtain business insights and actionable intelligence and knowledge, in order to support better business decision making and capture new business opportunities. This course will give students an understanding of the principles and practices of business intelligence and data analytics to support organizations in conducting their business in a competitive environment.

Learning Outcomes: Students, on successful completion of the course, will be able to

1. Explain the concepts characteristics of BI and data analytics
2. Describe multiple business problem/decision making domains requiring BI and data analytics
3. Apply BI and data analytic tools and technologies to develop BI applications
4. Integrate BI applications with other information systems as part of a business process
5. Define a BI strategy for an organization
6. Manage a BI project for an organization
7. Describe big data analytics and applications

Prerequisites: DMM

Course Outline:

I. Introduction to Business Intelligence
   1. BI Definition
   2. BI Concepts
   3. Business Intelligence, Analytics, and Data Science
   4. Business Intelligence to Support Decisions

II. Data Warehousing for BI
   1. DW design

2. Multidimensional data modelling and analysis
3. ETL process

III. Categories of Data analytics:

1. Descriptive Analytics
2. Predictive Analytics
3. Prescriptive Analytics

IV Descriptive Analytics

1. Descriptive Statistics
2. Business Performance Management
3. Data Visualization and Dashboard Design

V Predictive Analytics

1. Data Mining (Text Analytics and Text Mining, Web Analytics, Web Mining, and Social Analytics)
2. Predictive Modeling

VI Overview of Prescriptive Analytics

1. Optimization
2. Multi-Criteria Systems

VII Technical Aspects

1. BI Architecture
2. BI Tools and Technologies

VIII BI Applications

1. BI Maturity
2. BI Strategies
3. BI Project (case study)

IX Overview of Big Data

1. Big Data Analytics
2. Example of Big Data Applications

Laboratory Session(s): None

Learning Resources:

Text Book:

Business intelligence, analytics, and data science**,** by Ramesh Sharda; Dursun Delen; Efraim Turban, Pearson Publisher, 2018

Reference Books:

- Business Analytics (2nd Ed.) by James Evans, Pearson, 2017.
- Business Analysis for Business Intelligence (1st Ed) by Bert Brijs, Auerbach Publications, 2013.
- Business Intelligence Guidebook (1st Ed) by Rick Sherman, Morgan Kaufmann, 2014.
- Fundamentals of Business Intelligence by Wilfried Grossmann and Stefanie Rinderle-Ma**,** Springer, 2015.

Journals:

- Decision Support Systems
- International Journal of Business Intelligence and Data Mining
- International Journal of Business Intelligence Research Journal of Big Data

Teaching and Learning Methods:

1. Lecture
2. Assignment
3. Course Project
4. Real-world case studies
5. Self-learning

Evaluation Components

1. Exam
2. Assignment
3. Course Project

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

MACHINE LEARNING

Credits: 3-4

Objective: The course introduces students from a variety of science and engineering backgrounds to the fundamentals of machine learning and prepares them to perform R&D involving machine learning techniques and applications. Students learn to design, implement, and evaluate intelligent systems incorporating models learned from data.

Learning Outcomes: Students, on successful completion of the course, will be able to

1. Formulate a practical data analysis and prediction problem as a machine learning problem.
2. Identify the characteristics of the data set required for a particular machine learning problem.
3. Train and test supervised regression and classification models, unsupervised learning and density estimation models, and reinforcement learning models.
4. Integrate a trained machine learning model into an online software system.

Prerequisites: None

Course Outline:

I. Introduction to Machine Learning

II. Supervised Learning

   1. Linear regression, logistic regression, and generalized linear models
   2. Generative probabilistic models
   3. Convex optimization and quadratic programming
   4. Support vector machines
   5. Decision trees and ensemble models (should not be in AI)
   6. Non-parametric methods

III. Deep Learning

7. Perceptrons and inspiration from neuroscience
8. Multilayer neural networks and backpropagation
9. Optimization techniques, best practices, loss curve analysis

IV. <u>Learning Theory</u>

10. Bias-variance tradeoff
11. Regularization, model selection, and feature selection
12. Generalization bounds and VC dimension

V. <u>Unsupervised Learning</u>

13. Clustering: k-means, Gaussian mixture models
14. Principal components analysis
15. Independent components analysis
16. Autoencoders

VI. <u>Reinforcement Learning</u>

17. Markov decision processes and the Bellman equations
18. Value iteration, policy iteration
19. Q-learning

<u>Laboratory Session(s):</u>

1. Linear regression models
2. Logistic regression
3. Support vector classification
4. Decision trees
5. Single-layer and multi-layer neural networks
6. Multi-layer back-propagation, regularization, hyperparameter search
7. Model selection, feature selection
8. Clustering with k-means and GMMs
9. Principal components analysis and autoencoders
10. Value iteration and policy iteration
11. Q-learning
12. Deploying a machine learning model

Learning Resources:

Textbooks:

No designated textbook, but class notes and handouts will be provided.

Reference Books:

- Mitchell, T. (1997), *Machine Learning*, McGraw-Hill.
- Bishop, C. (2006), *Pattern Recognition and Machine Learning*, Springer. Goodfellow, I., Bengio, Y., and Courville, A. (2016), *Deep Learning*, MIT Press.
- Hastie, T., Tibshirani, R., and Friedman, J. (2016), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edition, Springer.
- Sutton, R.S. and Barto, A.G. (2018), *Reinforcement Learning: An Introduction*, 2nd Edition, MIT Press.

Journals and Magazines:

- *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*
- *Journal of Machine Learning Research (JMLR)*. Microtome

Others:

- Proceedings of the *Advances in Neural Information Processing Systems (NIPS)* conference
- Proceedings of the International Conference on Machine Learning (ICML).
- Lecture notes, posted online.

Teaching and Learning Methods:

1. **Use of online resources outside of class:** Students will be periodically assigned online video lectures prior to the face-to-face lecture.
2. Lectures
3. **In-class tutorials**: Tutorials on important data analysis and modeling tools will be given in class periodically.
4. **Laboratory sessions**: Students will be required to perform a series of exercises in data analysis and submit a lab report.
5. **Homework**: Several homework exercises requiring students to apply the knowledge acquired from lecture and discussion will be assigned and graded.

6. **Project**: Students will propose and execute a plan for a significant machine learning project in groups of 1-3. Students should formulate their data analysis problems independently under the guidance of the instructor, deploy a prototype, and make a formal present the results.

Evaluation Components

1. Term project outcome and presentation
2. Examinations
3. Homework
4. Lab reports

COMPUTER PROGRAMMING FOR DATA SCIENCE AND ARTIFICIAL INTELLIGENCE

Credits: 2

Objective: This course is a laboratory course that provides students with the computer programming background required for success in data science and artificial intelligence.

Learning Outcomes: Students, on successful completion of the course, will be able to

1. Prepare data for further analysis using data analytic tools
2. Manipulate data sets programmatically
3. Perform exploratory data analysis programmatically
4. Apply basic text processing techniques to unstructured data sets
5. Visualize data sets effectively
6. Perform basic statistical analyses programmatically
7. Build data-driven predictive models

Prerequisites: None

Course Outline:

I. Fundamentals
   1. Python programming
   2. The Python toolset

II. Working with data
   1. Numerical computation using numpy
   2. Data manipulation using pandas
   3. Exploratory data analysis
   4. Text processing with nltk

III. Data visualization
   1. Matplotlib
   2. Pandas
   3. Visdom

IV. Statistics

1. Random variables
2. Probability distributions
3. Hypothesis testing using scipy and statsmodels

V. <u>Machine learning tools</u>
1. Scikit-learn
2. Pytorch

<u>Laboratory Session(s)</u>: Each topic is a series of lab sessions.

<u>Learning Resources</u>:

<u>Textbooks</u>:

No specific textbook. Lab manuals and online resources will be used.

<u>Reference Books</u>:

Downey, A. (2014), *Think Stats*, 2nd edition, O'Reilly. McKinney, W. (2013), *Python for Data Analysis*, O'Reilly.

<u>Others</u>:

- Python tutorials available online
- Jupyter notebook tutorials availabl online
- Numpy tutorials available online
- Pandas tutorials available online
- Nltk tutorials available online
- Matplotlib tutorials available online
- Visdom tutorials available online
- Scikit-learn tutorials available online
- Pytorch tutorials available online

<u>Teaching and Learning Methods</u>:

1. **Use of online tutorials:** Students will make use of online tutorials for self-learning.
2. **Laboratory sessions**: Students will be required to perform a series of exercises and submit a lab report.

3. **Homework**: Several homework exercises requiring students to apply the knowledge acquired from lab and discussion will be assigned and graded.

Evaluation Components

1. Exam
2. Homework
3. Lab reports

Elective Courses

COMPUTATIONAL LINGUISTICS

Credits: 3

Objective: Students will understand problems and methods-computational linguistics and natural language processing. Students will learn the fundamentals of computational linguistics and its applications in text mining. Students will be able to apply the pre-processing and parsing methods for natural languages. Students will employ techniques and models for NLP problem scenarios, design, and implement solutions of NLP applications.

Learning Outcomes: Students, on successful completion of the course, will be able to

1. Explain the fundamentals of computational linguistics and natural language processing (NLP)
2. Apply state-of-the-art pre-processing and parsing methods for natural languages,
3. Describe and employ suitable machine learning techniques and models for NLP problem scenarios,
4. Design and implement NLP applications.

Prerequisites: AI, ML, Computer Programming for Data Science

Course Outline:

I. Introduction to computational linguistics and natural language processing

II. Pre-processing
   1. Tokenization and sentence splitting, morphology
   2. Regular Expressions, Edit Distance

III. Language processing
   1. POS tagging
   2. Sequence labelling and recurrent neural networks
   3. Sequence to sequence transformation

IV. Word Representations (vector semantics and embeddings)
   1. Vector space model and classification

2. Static word embeddings

V. Syntactic processing
1. Constituent parsing
2. Dependency parsing

VI. Semantic Analysis
1. Word sense disambiguation
2. Semantic role labeling

VII. Classification models
1. Naïve Bayes
2. Feedforward neural networks

VIII. Neural NLP and transfer learning

IX. Applications of Computational Linguistics
1. Information extraction (NER and Relation Extraction)
2. Sentiment analysis (Document level sentiment analysis and aspect based sentiment analysis)
3. Dialogue systems /Conversational agents
4. Machine translation

Learning Resources:

Textbooks / Reference Books:

- Textbook: Jurafsky, D., Martin, J. H., “Speech and Language Processing”, 3rd edition (online, 2019) https://web.stanford.edu/~jurafsky/slp3/
  One chapter from: Eisenstein, J., “Natural Language Processing” (online, 2018) https://github.com/jacobeisenstein/gt-nlp-class/tree/master/notes

Tutorials:

- Language processing with Spacy: https://realpython.com/natural-language-processing-spacy-python/
- Scikit-learn: Working with text data. https://scikit-learn.org/stable/tutorial/text_analytics/working_with_text_data.html
- CRFsuite: sequence labelling for Named entity recognition https://sklearn-crfsuite.readthedocs.io/en/latest/tutorial.html
- Word embeddings in Python with gensim: https://machinelearningmastery.com/develop-word-embeddings-python-gensim/

- BERT word embeddings in Huggingface: https://mccormickml.com/2019/05/14/BERT-word-embeddings-tutorial/
- GPT2-2 text generation tutorial: https://minimaxir.com/2019/09/howto-gpt2/

Assignments:

- Assignment_1_text_categorization.pdf
- Assignment_2_sequence_labelling.pdf
- Assignment_3_sentiment_analysis.pdf

Teaching and Learning Methods:

1. *Direct instruction-based teaching* for the lectures using visual aid via slides
2. *Self-learning* method for NLP models using tutorials
3. *Inquiry based teaching* for the labs through one medium scale project and assignments

Evaluation Components

1. Examinations
2. Assignments
3. Project on Computational Linguistics

COMPUTER VISION

Credits: 3

Objective: To introduce the concepts of computer vision with emphasis on state-of-the-art methods used in vision applications.

Learning Outcomes: Students, on successful completion of the course, will be able to

1. Explain key concepts of computer vision.
2. Extract discriminative features from image/video data and use them for pattern classification.
3. Analyse, examine, and evaluate existing practical computer vision systems.
4. Apply computer vision algorithms from standard libraries and tools to build prototype computer vision systems for real scenarios.

Prerequisites: AI and ML

Course Outline:

I. Introduction to Computer Vision
   1. Mathematical foundation for Computer Vision
   2. Geometry of image formation
   3. Image filtering and Edge detection
   4. Image segmentation

II. Feature detection and matching
   1. Feature detection methods
   2. Feature description methods
   3. Feature matching

III. Object Recognition
   1. Image classification
   2. Bag of words
   3. Convolutional neural networks
   4. Generative adversarial networks and recurrent neural networks
   5. Performance evaluation

IV. Motion Analysis and Tracking

1. Camera models
2. Two-view geometry
3. Stereo
4. Optical flow
5. Structure from motion

V. Case Studies and Applications of Computer Vision

1. Large scale image search and feature indexing
2. Trajectory analysis
3. Image caption generation
4. Action recognition
5. Other state of the art applications of computer vision

Laboratory Session(s): 30 hours of laboratory sessions on Open CV, Python deep learning frameworks

Learning Resources:

Reference Books

- Szeliski, R. (2010). *Computer Vision: Algorithms and Applications*. Springer.
- Forsyth, D.A. and Ponce, J. (2011). *Computer Vision: A Modern Approach*. 2nd Ed. Prentice Hall.
- Goodfellow, I., Bengio, Y. and Courville, A. (2016). Deep Learning. MIT Press

Recommended Readings

Hartley, R and Zisserman, A. (2004), Multiple View Geometry in Computer Vision. 2nd Ed. Cambridge University Press

Teaching and Learning Methods:

1. *Direct instruction*-based teaching for the lectures using visual aid via slides.
2. *Self-learning* methods through activities facilitated by the teacher.
3. **Use of online resources outside of class:** Students will be periodically assigned online video lectures prior to the face-to-face lecture.
4. **In-class tutorials**: Tutorials on cloud computing.

5. **Laboratory sessions**: Students will be required to perform a series of exercises and submit a lab report.
6. **Homework**: Several homework exercises requiring students to apply the knowledge acquired from lecture and discussion will be assigned and graded.
7. **Project**: Students will propose and execute a plan for a computer vision project in groups of 2-3.

Evaluation Components

1. Exam
2. Assignments
3. Projects

DISTRIBUTED SYSTEMS

Credits: 3

Objective: The course introduces the concepts of distributed systems, cloud computing, and blockchain. Students learn to create required distributed infrastructure and ecosystems for DS&AI applications. Students earn skills on deployment, monitoring, and management of distributed systems.

Learning Outcomes: Students, on successful completion of the course, will be able to

1. Explain the main concepts of distributed systems, cloud computing, and blockchain,
2. Setup distributed environment for DS&AI applications,
3. Utilize distributed file systems,
4. Create the process pipeline and deploy cloud services,
5. Monitor and management the usage of network resources,
6. Implement applications with blockchain/smart contracts.

Prerequisites: None

Course Outline:

I. Distributed systems
    1. Network topologies, protocols, and synchronization
    2. Replication, Consistency, and Fault Tolerance
    3. Security
    4. Types of distributed file systems (DFS)
    5. Distributed Web-Based Systems
    6. Distributed Coordination-Based Systems
    7. Content-delivery network (CDN) and Software-defined network (SDN)

II. Cloud Computing
    1. Characteristics of cloud computing
    2. Virtualization and containers
    3. Service models (SaaS, IaaS, and PaaS)
    4. Ecosystems and container-orchestration systems

5. Recent trends in cloud computing

III. Blockchain

1. Distributed ledgers and blockchain
2. Blockchain Network and Data Structures
3. Blockchain as a platform
4. Security vulnerability in smart contracts
5. Consensus
6. Economics and Review
7. BFT Protocols
8. BFT Anonymity and Privacy
9. Secure Multiparty Computation and Applications
10. Networks and Applications
11. Cryptography and hash algorithms

Exercises/Laboratory Session(s):

1. Distributed environment
2. Distribute file systems
3. Virtualization and containers
4. Container-orchestration systems
5. Blockchain/smart contracts

Learning Resources:

Textbooks:

- Kshemkalyani, A. D. (2011). *Distributed Computing: Principles, Algorithms, and Systems*, Cambridge University Press.
- Kirk, D. B. and Hwu, W. W. (2016). *Programming Massively Parallel Processors: A Hands-on Approach*, 3rd edition, Morgan Kaufmann.
- Rafaels, R. (2018). *Cloud Computing: 2018*, 2nd edition, CreateSpace Independent Publishing Platform.

Journals and Magazines:

IEEE Transactions on Services Computing

Others:

- Lecture notes posted online.
- Official online documentations, e.g.,
  - Kubernetes documentation (https://kubernetes.io/docs/home/)
  - Docker documentation (https://docs.docker.com/)

Teaching and Learning Methods:

1. **Use of online resources outside of class:** Students will be periodically assigned online video lectures prior to the face-to-face lecture.
2. **Lectures:** Direct instruction instruction-based teaching for the lectures using visual aid via slides.
3. **In-class tutorials**: Tutorials on cloud computing.
4. **Laboratory sessions**: Students will be required to perform a series of exercises.
5. **Homework**: Several homework exercises requiring students to apply the knowledge acquired from lecture and discussion will be assigned and graded.
6. **Project**: Students will propose and execute a plan for blockchain/smart contracts project in groups of 2-3.

Evaluation Components

1. Examinations
2. Homework
3. Lab evaluations
4. Project

HUMAN COMPUTER INTERACTION AND INFORMATION VISUALISATION

Credits: 3

Objective: Students should understand the principles, processes and techniques for design, implementation and evaluation of interactive systems to maximize usability and to enhance user experience of data-driven systems. Students would learn the methods and techniques to present information to enhance the understanding of data.

Learning Outcomes: Students, on successful completion of the course, will be able to

1. Explain capabilities of both humans and computers and the theoretical foundation of human computer interaction (HCI)
2. Adopt the process of design thinking for development of interactive systems
3. Employ tools in HCI for implementation of systems with maximized usability and enhanced user experience
4. Explain the fundamentals of information visualization
5. Summarize dynamic, real-time and spatial datasets across categories, space, and time through visualization tools.

Prerequisites: DMM

Course Outline:

I. Introduction to HCI
   1. Humans and computers
   2. Interaction

II. Design process
   1. Interaction and design basics
   2. User centred design
   3. Design rules
   4. Implementation
   5. Evaluation techniques
   6. Design thinking

III. Models and theories of HCI
   1. Cognitive models
   2. Communication and collaboration models
   3. Task analysis
   4. Models of the system
   5. Modelling rich interaction

IV. Introduction information visualization
   1. Theories of data graphics
   2. Static and moving patterns
   3. Visual objects and data objects

V. Data and visualization design, implementation and evaluation
   1. Data types and visualizations
   2. Space Perception and the Display of Data in Space
   3. Interacting with Visualizations
   4. Visualization design, implementation and evaluation

Laboratory Session(s): 30 hours of laboratory sessions on prolog or R or Microsoft Power BI or Tabulae etc. using real data and creating visualizations.

Learning Resources:

Reference Books:

- Alan Dix et. al. (2003), *Human Computer Interaction*, 3rd edition, Pearson.
- Colin Ware (2012), *Information Visualization: Perception for Design*, 3rd edition, Morga Kaufmann.

Additional Reading:

- Cole Nussbaumer Knaflic (2015), *Storytelling with Data: A Data Visualization Guide for Business Professionals*, 1st edition, Wiley.
- Steve Wexler, Jeffrey Shaffer and Andy Cotgreave (2017), *The Big Book of Dashboards: Visualizing Your Data Using Real-World Business Scenarios,* 1st edition, Wiley.

Teaching and Learning Methods:

1. *Direct instruction* based teaching for the lectures using visual aid via slides.

2. *Self-learning* methods through activities facilitated by the teacher.
3. *Inquire based* teaching for the labs through multiple small-scale projects and assignments related to section II, III and V in the course outline.

Evaluation Components

1. Exam
2. Assignments
3. Projects

KNOWLEDGE REPRESENTATION

Credits: 3

Objective: Introduce the students to the field of knowledge representation, with the goal of reasoning about knowledge. The students will be exposed to specialized knowledge representations stemming from applications in different domains, such as, semantic web and cognitive robotics.

Learning Outcomes: Students, on successful completion of the course, will be able to

1. Use logical formalisms to effectively describe knowledge, belief, events, and situations
2. Identify the components of nonmonotonic reasoning and its usefulness as representation mechanism for knowledge systems
3. Design real world knowledge-based systems

Prerequisites: Artificial Intelligence.

Course Outline:

I. Knowledge Representation and Reasoning
   1. Ontologies
   2. Conceptual Graphs
   3. Linked Data and Semantic Web
   4. Description Logics and UML
   5. Nonmonotonic Reasoning,
   6. Answer Sets
   7. Rule-based, Model-based and Case-based Reasoning

II. Classes of Knowledge and Specialized Representations
   1. Reasoning about Knowledge and Belief
   2. Reasoning about Actions, Events, and time
   3. Situation Calculus
   4. Event Calculus and Transaction Logic

III. Knowledge Representation in Applications
   1. Semantic Web

2. Cognitive Robotics
3. Knowledge Engineering

Practical Sessions: The lectures are accompanied by practical session where the students work in small groups to solve real problems based on the knowledge acquired during the lectures.

Learning Resources:

Reference Books:

- van Harmelen, F., Lifschitz, V., and Porter, B., editors (2007), *Handbook of Knowledge Representation,* 1st edition, Elsevier.
- Heath, T. and Bizer, C. (2011), *Linked Data: Evolving the Web into a Global Data Space*, 1st edition. Synthesis Lectures on the Semantic Web: Theory and Technology, Morgan & Claypool.

Teaching and Learning Methods:

1. *Direct instruction* based teaching for the lectures using visual aid via slides
2. *Self-learning* method for in-depth knowledge of specialized topics
3. *Inquire based* teaching for the practical sessions

Evaluation Components

1. Exam
2. Group assignments

MULTICRITERIA OPTIMIZATION AND DECISION ANALYSIS

Credits: 3

Objective: This course will give students an understanding of the decision making process and multicriteria decision analysis methods and optimization processes for finding optimal solutions to problems with multiple decision alternatives and conflicting objectives.

Learning Outcomes: Students, on successful completion of the course, will be able t

1. Describe the decision making processes typically used by organizations.
2. Formulate a decision making scenario as a multicriteria decision analysis problem.
3. Identify and formulate different types of mathematical programming problems including formulations with constraints and multiple objectives.
4. Analytically solve simple Pareto optimization problems that are special cases for the application of Karush-Kuhn-Tucker conditions and the Lagrange multiplier theorem.
5. Apply methods of multicriteria optimization and decision analysis to real world problem domains.

Prerequisites: None.

Course Outline:

I. Introduction to MODA
   1. Topic and History

II. Problem Classification
   1. Linear and Mathematical Programming
   2. Complexity issues

III. Multicriteria decision analysis
   1. A priori multicriteria decision making
   2. A posteriori multicriteria decision making

IV. Orders and Pareto Dominance
   1. Pareto Dominance orders in Multiobjective Optimization
   2. Multicriteria Landscape Analysis

3. Linear approximation of a nonlinear function

V. Maxima of a Set of Vectors
   1. Techniques and computational complexity

VI. Evolutionary Multiobjective Optimization
   1. Evolutionary Optimization Algorithms
   2. Few algorithms: NSGA II, SMS-EMOA, and MOCOPS

VII. Applications (examples)
   1. Health care
   2. Project management
   3. Water management

VIII. Tools and software
   1. Expert Choice
   2. Decisionarium

Laboratory Session(s): Tutorials are provided, upon request.

Learning Resources:

Textbooks:

Ishizaka, A., and Nemery, P. (2013). *Multi-Criteria Decision Analysis:Methods and Software,* John Wiley & Sons.

Reference Books:

- Kaliszewski, I., Miroforidis, J., and Podkopaev, D. (2016), *Multiple Criteria Decision Making by Multiobjective Optimization: A Toolbox*, Springer.
- Ehrgott, M. (2000), *Multicriteria Optimization*, Springer.

Journals and Magazines:

Journal of Multi-Criteria Decision Making, Wiley.

Others:

A.H. Deutz, M.T. M. Emmerich (2018) A tutorial on multiobjective optimization: fundamentals and evolutionary methods. *Natural Computing* 17:585–609.

Teaching and Learning Methods:

1. Lecture
2. Assignments
3. Course Project
4. Real-world case studies
5. Self-learning

Evaluation Components

1. Exam
2. Assignments
3. Course Project

NATURE-INSPIRED COMPUTING

Credits: 3

Objective: Introduce the students to the field of nature-inspired metaheuristic methods for search and optimization, including the latest trends in nature-inspired algorithms and other forms of natural computing. The students will be exposed not only to paradigms of nature-inspired metaheuristic methods (originating from, for example, biology, living thing behavior and natural phenomena), but also to their applications.

Learning Outcomes: Students, on successful completion of the course, will be able to

1. demonstrate fundamental insights of nature-inspired computation;
2. implement nature-inspired methods into concrete algorithms;
3. apply nature-inspired algorithms to some search and optimization applications.

Prerequisites: none.

Course Outline:

I. Bio-Inspired Algorithms
1. Genetic Algorithm
2. Clonal Selection Algorithm

II. Living Thing Behaviours
1. Particle Swarm Optimization
2. Jaya algorithm
3. Grey Wolf Optimizer
4. Biogeography-Based Optimization

III. Other Paradigms
1. Differential Evolution
2. Cuckoo Search
3. Simulated Annealing
4. Evolutionary Strategies

IV. <u>Strategies for Empowering the Search</u>
1. Chaotic-Based Approach
2. Oppositional-Based Approach
3. Elitism-Based Approach
4. Multi-species Approach
5. Cooperative Approach
6. Hybridizing Approach

V. <u>Nature-Inspired Computing in Applications</u>
1. Handle the Constraints of Problem
2. Machine Learning Construction
3. Cluster Data
4. Optimize Engineering Problem
5. Handle Multi-objective

<u>Practical Sessions</u>: The lectures are accompanied by practical sessions where students work in small groups to learn algorithms or solve problems.

<u>Learning Resources</u>:

<u>Textbooks</u>: No designated textbook, but class notes and handouts will be provided.

<u>Reference Books</u>:

- Xin-She Yang (2014), Nature-Inspired Optimization Algorithms (Elsevier Insights) 1st Edition, Elsevier Insights.
- David E. Goldberg (1989), Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley Professional; 1 edition.
- Dan Simon (2013), Evolutionary Optimization Algorithms: Biologically-Inspired and Population-Based Approaches to Computer Intelligence, John Wiley & Sons.
- Ravipudi Venkata Rao (2019), Jaya: An advanced Optimization Algorithm and its Engineering Applications, Springer International Publishing AG, part of Springer Nature 2019.

<u>Journals and Magazines:</u>

- *IEEE Transaction on Evolutionary Computing*

- *Swarm and Evolutionary Computation*

Others:

- Kirkpatrick, S.; Gelatt Jr, C. D.; Vecchi, M. P. (1983). "Optimization by Simulated Annealing". Science. 220 (4598): 671–680. Bibcode:1983Sci...220..671K. CiteSeerX 10.1.1.123.7607. doi:10.1126/science.220.4598.671. JSTOR 1690046. PMID 17813860
- L. N. de Castro and F. J. Von Zuben, "Learning and optimization using the clonal selection principle," in *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 3, pp. 239-251, June 2002. doi: 10.1109/TEVC.2002.1011539
- *Seyedali Mirjalili, Seyed Mohammad Mirjalili, Andrew Lewis, Grey Wolf Optimizer, Advances in Engineering Software, Volume 69, 2014, Pages 46-61, https://doi.org/10.1016/j.advengsoft.2013.12.007.*
- D. Simon, "Biogeography-Based Optimization," in *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 6, pp. 702-713, Dec. 2008. doi: 10.1109/TEVC.2008.919004

Teaching and Learning Methods:

1. *Direct instruction*-based teaching for the lectures using visual aid via slides and program codes
2. *Self-learning* method for in-depth knowledge of specialized topics.
3. *Inquire-based* method for the practical sessions

Evaluation Components:

1. Exam
2. Group assignments
3. Group project

RECENT TRENDS IN MACHINE LEARNING

Credits: 3

Objective: The course builds on the content of Machine Learning, providing students with a deeper understanding of machine learning techniques and a wider variety of extant learning models. Students will be prepared to develop advanced machine learning applications and perform research at a state-of-the-art level.

Learning Outcomes: Students, on successful completion of the course, will be able to

1. Design, train, test, and deploy modern convolutional neural networks (CNNs).
2. Utilize the principles of adversarial learning to increase the robustness of a machine learning model.
3. Design, train, test, and deploy generative adversarial networks (GANs).
4. Utilize recurrent neural networks (RNNs) to model and predict time series.
5. Utilize deep neural networks to solve difficult tabula rasa reinforcement learning problems.
6. Apply state-of-the-art machine learning methods to solve problems in speech processing, speech synthesis, natural language understanding, natural language synthesis, computer vision, and intelligent agent design.

Prerequisites: None

Course Outline:

I. Overview of modern machine learning methods

II. Convolutional neural networks
   1. Fundamentals
   2. Inception modules
   3. Residual layers
   4. Squeeze and excitation
   5. Detection models
   6. Semantic segmentation models
   7. Instance-aware segmentation models

III. Deep Belief Networks

1. Belief Nets
2. Restricted Boltzmann Machines
3. Semi-restricted Boltzmann Machines
4. Higher order Boltzmann machines

IV. Transfer learning
1. Inductive transfer learning
2. Transductive transfer learning
3. Unsupervised transfer learning

V. Automatic learning
1. Automated feature engineering
2. Automated model selection
3. Automated optimization algorithm selection

VI. Deep unsupervised learning
1. Generative adversarial networks (GANs)
2. Cycle GANs
3. Wasserstein GANs
4. Variational autoencoders

VII. Practical techniques for deep learning models
1. Weight initialization
2. Dropout
3. Adam optimization
4. Batch normalization

VIII. Time series processing
1. Hidden Markov models (HMMs)
2. Recurrent neural networks (RNNs) and backpropagation through time
3. Word embedding for natural language processing
4. Long short term memory (LSTM) units
5. Gated recurrent units (GRUs)
6. Attention mechanisms for RNNs

IX. (Deep) Reinforcement learning
1. Policy gradients
2. Actor/critic methods
3. Imitation learning
4. Exploration/exploitation
5. Meta learning

6. Monte Carlo methods

X. Applications
    1. Speech recognition
    2. Speech synthesis
    3. Conversational agents
    4. Recommendation systems
    5. Anomaly detection
    6. Computer vision systems

Laboratory Session(s):

1. Preparing the environment for machine learning tools
2. CNNs and residual layers
3. Generative adversarial networks (GANs)
4. Deep learning techniques
5. Introductory time series processing
6. Time series processing with LSTMs and GRUs
7. Deep reinforcement learning
8. Deep speech recognition
9. Recommendation systems
10. Anomaly detection
11. Computer vision

Learning Resources:

Textbooks: No designated textbook. Emphasis is on recent papers in major machine learning conferences. Class notes and handouts will be provided.

Reference Books:

- Goodfellow, I., Bengio, Y., and Courville, A. (2016), *Deep Learning*, MIT Press.
- Sutton, R.S. and Barto, A.G. (2018), *Reinforcement Learning: An Introduction*, 2nd edition, MIT Press.

Journals and Magazines:

- *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI).* IEEE
- *Journal of Machine Learning Research (JMLR)*. Microtome

Others:

- Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). IEEE
- Proceedings of the *Advances in Neural Information Processing Systems (NeurIPS)* conference. Neural Information Systems Foundation, Inc.
- Proceedings of the *International Conference on Machine Learning (ICML)*. International Machine Learning Society.
- Lecture notes: posted online.

Teaching and Learning Methods:

1. **Use of online resources outside of class:** Students will be periodically assigned online video lectures prior to the face-to-face lecture.
2. **Lectures**
3. **In-class tutorials**: Tutorials on important data analysis and modeling tools will be given in class periodically.
4. **Laboratory sessions**: Students will be required to perform a series of exercises in data analysis and submit a lab report.
5. **Homework**: Several homework exercises requiring students to apply the knowledge acquired from lecture and discussion will be assigned and graded.
6. **Project**: Students will propose and execute a plan for a significant machine learning project in groups of 1-3.

Evaluation Components

1. Term project
2. Midterm examination
3. Final examination
4. Homework
5. Lab reports

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

SOFTWARE DEVELOPMENT AND PROJECT MANAGEMENT FOR DS & AI

Credits: 3

Objective: The course emphasizes on modern and important software development, software process~~,~~ and project management. Student will tailor the software development process and project management for DS&AI projects, including planning, iterative development, test driven development, continuous integration/continuous delivery, versioning, and deliverables. Students learn to apply knowledge to the problems in DS&AI domains.

Learning Outcomes: Students, on successful completion of the course, will be able to

1. Explain the importance of software development and project management,
2. Explain how model-driven development works in a DevOps and agile environments,
3. Create model and data versioning,
4. Apply the principles of project management to DS & AI project.

Prerequisites: None

Course Outline:

I. Software Development and Software Process
   1. Introduction to Modern Software Process
      a) Agile practices and frameworks
      b) Model-driven development
   2. Test-Driven Development (TDD)
   3. Test Automation
   4. Continuous Integration/Continuous Delivery (CI/CD)
   5. Configuration Management
   6. DevOps

II. Project Management
   1. Project Integration Management
   2. Project Scope Management

3. Project Time Management
4. Project Cost Management
5. Project Quality Management
6. Project Human Resources Management
7. Project Communications Management
8. Project Risk Management
9. Project Procurement Management
10. Project Stakeholder Management

III. DS & AI project management

Laboratory Session(s): None

Learning Resources:

Textbook:

Project Management Institute. (2017). *A Guide to the Project Management Body of Knowledge (Pmbok Guide)*, 6th edition, The Stationery Office Ltd.

Reference Books:

- Beck, K. and Andres, C. (2004). *Extreme Programming Explained: Embrace Change: Embracing Change*, 2nd Edition, Addison-Wesley Professional.
- Forsgren, N., Humble, J., and Kim, G. (2018). *Accelerate: The Science of Lean Software and Devops: Building and Scaling High Performing Technology Organizations*, 1st Edition, IT Revolution Press.
- Rubin, K.S. (2012). *Essential Scrum: A Practical Guide to the Most Popular Agile Process (Addison-Wesley Signature): A Practical Guide To The Most Popular Agile Process (Addison-Wesley Signature Series (Cohn))*, 1st edition, Addison-Wesley Professional.
- Humble, J. and Farley, D. (2010), *Continuous Delivery: Reliable Software Releases through Build, Test, and Deployment Automation (Addison-Wesley Signature Series (Fowler))*, 1st edition, Addison-Wesley Professional.

Others:

Lecture notes, posted online.

Teaching and Learning Methods:

1. **Use of online resources outside of class:** Students will be periodically assigned online video lectures prior to the face-to-face lecture.
2. **Lectures**
3. **In-class tutorials**: Tutorials on important tools will be given in class.
4. Individual Projects: small individual projects that require students to apply the knowledge acquired from lecture and discussion will be assigned and graded.
5. **Group Project**: Students will propose and execute a plan for an application of DS&AI project in groups of 3-4 people.

Evaluation Components

1. Individual Projects
2. Group project
3. Examinations

SPATIAL-TEMPORAL DATA ANALYSIS

Credits: 3

Objective:

Students should understand problems, methods, algorithms, and novel computational techniques in the analysis of spatio-temporal databases. Students will apply these understanding in spatio-temporal data projects.

Learning Outcomes: Students, on successful completion of the course, will be able to

1. Explain the problems and methods (minimum methods are clustering and predictive learning) on the spatio-temporal data mining
2. Apply modeling skill for realizing spatio-temporal data projects
3. Apply integration skill for realizing spatio-temporal data projects
4. Apply visualization skill for realizing spatio-temporal data projects

Prerequisites:

1. Machine Learning
2. Data Modeling and Management
3. Programming for Data Science and Artificial Intelligence

Course Outline:

I. Description on Spatio-Temporal Data
   1. Examples on real case applications on Spatio-Temporal Data
   2. Definition and Properties of Spatio-Temporal Data
   3. Defining Features for Spatio-Temporal Data, including the data instance
   4. Data Similarity

II. Modeling on Spatio-Temporal Data
   1. Clustering on Spatio-Temporal Data
   2. Classification on Spatio-Temporal Data
   3. Predictive learning on Spatio-Temporal Data

III. Integration on Spatio-Temporal Data Projects
   1. Capturing Time Series Data

2. Capturing Spatial Data
3. Pre-processing on Time Series Data (time and frequency domain)
4. Pre-processing on Spatial Data
5. Data Management for the Spatio-Temporal Data

IV. Visualization on Spatio-Temporal Data
1. Visual mapping
2. Visualizing geo-spatial data
3. Visualizing spatio-temporal data

V. Applications of Spatio-Temporal Data Analysis

Laboratory Session(s): A list of specific lab sessions

Learning Resources:

Textbook:

Cressie, N., Wikle, C. K., “Statistics for Spatio-Temporal Data”, John Wiley & Sons, 1st edition, 2015

Reference Books:

- Zheng, Y. (2019). *Urban computing*. MIT Press.
- Hsu, W., Lee, M. L., & Wang, J. (2008). *Temporal and spatio-temporal data mining*. IGI Pub..

Others:

- Atluri, G., Karpatne, A., & Kumar, V. (2018). Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys (CSUR)*, *51*(4), 1-41.
- Zheng, Y., Capra, L., Wolfson, O., & Yang, H. (2014). Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *5*(3), 1-55.

Teaching and Learning Methods:

1. Lectures
2. Assignments
3. Laboratory sessions
4. Group Project
5. Self-learning

Evaluation Components

1. Exam
2. Assignment
3. Project on Spatio-Temporal Data Mining

SOCIAL NETWORK ANALYSIS (SNA)

Credits: 3

Objective: With the growth and popularity of social network, the analysis of large datasets of network is becoming more important. This course will give the students an understanding of methods of social networks analysis and applications of social network analysis.

Learning Outcomes: Students, on successful completion of the course, will be able to

1. Describe different algorithms and methods of social network analysis.
2. Know the evolution of social network
3. Use SNA for studying the social behaviour or social structure to making decisions.
4. Develop applications of SNA

Prerequisites: none

Course Outline:

I. Social network mining:
    1. An Overview of methods for social network analysis (SNA): for getting foundation from core subjects of data science to face challenging in the world
    2. Discovering sets of key players in social networks and data socially similar
    3. Clustering and extracting social network form web.
    4. Exploratory SNA.
    5. Information flow in system interacting agents, influence from local to global.

II. Social network evolution:

    Network: theory and mechanism – from classical SNA to possibilities change.
    1. Layout of the networks: for drawing and interpreting.
    2. Nature and innovation dissemination of information in the network: internet, web, etc.

3. Interdisciplinary Matchmaking – choosing collaborators by skill, acquaintance, and trust/information retrieval.
4. Extended generalized block-modelling for compound communities and external actors
5. Analysing collaborations through content-based social networks.

III. Social network applications (examples):

1. Social network for radio communication.
2. Social networks web services.
3. Customer relationship management (CRM)
4. Crime analysis and control

IV. SNA Software and Tools

1. AllegroGraph
2. GraphStream
3. Graph-Tools

Laboratory Session(s): A list of specific lab sessions

Learning Resources:

Textbook:

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge university press.

Reference Books:

- Scott, J. (2017). Popularity, mediation and exclusion. *Social Network Analysis*, 95-112.
- Mika, P. (2007). The Semantic Web. *Social Networks and the Semantic Web*, 3-26.
- Yang, S., Keller, F. B., & Zheng, L. (2016). *Social network analysis: Methods and examples*. Sage Publications.

Teaching and Learning Methods:

1. Lectures
2. Assignment
3. Course Project
4. Real-world case studies
5. Self-leaning

Evaluation Components

1. Exam
2. Assignment
3. Course Project

Professional Courses

REAL WORLD BIG DATA ENGINEERING

Description: Data engineers work closely with data scientists and are largely in charge of architecting solutions for data scientists. Data engineers are responsible mainly for tasked with transforming data into a format that can be easily analyzed. They do this by developing, maintaining, and testing infrastructures for data generation. However, data engineering also need to understand how to optimize data retrieval and how to develop data visualization such as dashboard, reports and other visualization for stakeholder's demand. The huge scale of data often has various and multiple data analysts or scientists to help understand data. In addition, data engineers are responsible for finding trends in data sets including required a significant set of technical skills, a deep knowledge of SQL database design and multiple programming languages. In addition, the data engineer also need communication skills to work across departments to understand what meet the business goal from the company's large datasets.

Training Duration: 24 hours (3 days)

Target Audience:

- Business Analyst
- Database Administrator
- IT Architect/ Data Architect
- Developer Operation
- Full-stack Developer

Required Skills:

- Having knowledge of SQL language
- Having knowledge of operating systems command.
- Having experience in programming Python, R language.
- At least one year of IT experience would be preferred.

Objective:

- To identify and describe the data engineering concepts why data engineering is very importance.
- To identify and describe the process of data engineering and framework of data engineering tools such as Hadoop, Spark and so on.
- To get hands-on experience on HDFS via Hadoop file system command and sharing resources with cluster server.
- To get hands-on experience on applied data engineering for business cases.
- To perform data engineering skills for data analytics-life cycle.

Learning Outcomes: On successful completion of the training course, the training professionals will be able to:

1. Able to identify and describe the concepts and the importance of data engineering.
2. Able to identify and describe the process of data engineering and component of tools such as Hadoop, Spark so on.
3. Get experience on the process of Hadoop file system.
4. Apply problem solving to the huge scale of dataset in business cases.
5. Enhance skills of data engineering in real business use cases.

Training Course Outline

Day1:

- Introduction to Data Engineering? Why data engineering importance?
- Use case I
- Use case II
- The component of hadoop ecosystem.
- Preparation and setting up of big data environment and infrastructure.

Day2:

- Practice to install of Hadoop (Step by Step)

- Hadoop ecosystem I: Practice to use distributed file system using HDFS, Flume, Hive, SparkR, SySpark, Spark SQL.
- Hadoop ecosystem II: Practice to use Pig, HBase, Mahout, and Sqoop.

Day3:

- Apply to real business use cases with data engineering gathering ML toolbox to data engineering simply.

Application I:

- Assignment and presentation by grouping
- Applied to real business use cases in data engineering gathering to data visualization tools.

Application II:

- Assignment and presentation by grouping

Assessment:

- Pre-test and Post-test
- Assignment and Presentations

NOSQL DATABASE DESIGN AND DEVELOPMENT

Description: This training course emphasizes on emerging databases and technologies suitable for managing different types and characteristics of data. The course covers major categories of NoSQL data management systems including key-value, document, columnar and graph, which are increasing in popularity due to the growth of non-relational data and can solve problems relational databases cannot handle. The course explores the importance and applications of NoSQL databases, the classification of the databases and how to use them in a real-world scenarios.

The course will adopt a technical hands-on approach with practical tools, examples, exercises and business use cases to design, create databases, load and query data.

Training Duration: 24 hours (3 days)

Target Audience:

- Business Analysts
- Database System Specialists
- IT Specialists

Required Skills: Knowledge of relational database modeling and SQL.

Learning Outcomes: On successful completion of the training course, the training professionals will be able to:

1. Describe the differences between relational and NoSQL databases.
2. Design and organize various types of data (unstructured, semi-structured and structured) using NoSQL databases.
3. Identify, implement and perform frequent data operations (CRUD: create, read, update and delete) on NoSQL databases.

Training Course Outline:

Day1

- Introduction to the training course
- Application and Business Use Case I
- Application and Business Use Case II
- Recall: Relational Database Concepts and SQL

- NoSQL Database Concepts

Day2:

- Key-value Databases
- Document Databases
- Hands-on Sessions and Case Studies

Day3

- Columnar Databases
- Graph Databases
- Hands-on Sessions and Case Studies
- Conclusions

MACHINE LEARNING FOR PROFESSIONALS

Description: This professional training course emphasizes on machine learning techniques and applications for finding interesting patterns / information from large amount of data. Participants will learn to design, implement, and evaluate intelligent systems incorporating models learned from large data.

Training Period: 24 hours (3 days)

Target Audience: people interested in data sciences, beginner in data analyst, fresh-graduates audience in statistics, mathematics, or CS.

Required Skills: Programming using Python

Learning Outcomes: On successful completion of the training course, the training professionals will be able to:

1. Formulate practical data analysis and understand basic machine learning techniques
2. Identify the characteristics of the data set given a specific machine learning problem
3. Conduct training and testing in classification/ supervised learning
4. Integrate a trained machine learning model into an online software system.

Training Course Outline:

Day1:

- What is machine learning
- Python ML Libraries: Numpy, Pandas, Mapplotlib, Scikitlearn
- Steps in typical machine learning projects
- Fitting model to data
- Optimising cost function
- Handling, cleaning, and preparing data
- Selecting and engineering features
- Selecting model and tuning hyper-parameter
- Challenges in Machine Learning: underfitting and overfitting problems
- Feature extraction

Day2:

- Training, validation, and testing datasets
- Logistic Regression and softmax
- K-Nearest Neighbors, SVM, Decision Tree

Day3:

- Dimensionality Reduction: Principal components analysis
- K-Means and K-Medians

Session(s): This professional course will fully provide the hands-on.

Learning Resources:

Textbooks:

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media.