Erasmus Mundus Joint Master Degree

Big Data Management and Analytics

Detailed Course Description

Academic Year 2022-2023

Contents

Université libre de Bruxelles (ULB) **3**
Advanced Databases (ADB) . . . 3
Database Systems Architecture (DBSA) . . . 4
Data Warehouses (DW) . . . 5
Management of Data Science and Business Workflows (WM) . . . 7
Data Mining (DM) . . . 9

Universitat Politècnica de Catalunya (UPC) **11**
Big Data Management (BDM) . . . 11
Semantic Data Management (SDM) . . . 13
Machine Learning (ML) . . . 15
Viability of Business Projects (VBP) . . . 17
Big Data Seminar (BDS) . . . 19
Debates on Ethics of Big Data (DEBD) . . . 20

Universitá degli Studi di Padova (UniPD) **21**
Statistical Learning (StatLearn) . . . 21
Deep Learning and Human Data Analytics (DeepLearn) . . . 22
Time-Series Analysis for Business Economic and Financial Data (TimeSeries) . . . 24

Eindhoven University of Technology (TU/e) **26**
Foundations of Process Mining (2AMI10) . . . 26
Longitudinal Data Analysis (2AMS10) . . . 28
Responsible Data Challenge (2AMR10) . . . 29
Advanced Process Mining (2AMI20) . . . 30
Applications of Data Science for Software Engineering (2IMP40) . . . 32
Seminar Process Analytics (2IMI00) . . . 33

CentraleSupélec (CS) **35**
Decision Modeling (DeM) . . . 35
Advanced Machine Learning (ML) . . . 36
Visual Analytics (VA) . . . 37
Massive Graph Management & Analytics (MGMA) . . . 38
Big Data Research Project (BDRP) . . . 39
Business Innovation Management (BIM) . . . 40

University: Université Libre de Bruxelles (ULB)
Department: École polytechnique de Bruxelles
Course ID: ADB (INFO-H-415)
Course name: Advanced Databases
Name and email address of the instructors: Esteban Zimányi (ezimanyi@ulb.ac.be)
Web page of the course: http://cs.ulb.ac.be/public/teaching/infoh415
Semester: 1
Number of ECTS: 5

Course breakdown and hours:

- Lectures: 24h.
- Exercises: 24h.
- Projects: 12h.

Goals:
Today, databases are moving away from typical management applications, and address new application areas. For this, databases must consider (1) recent developments in computer technology, as the object paradigm and distribution, and (2) management of new data types such as spatial or temporal data. This course introduces the concepts and techniques of some innovative database applications

Learning outcomes:
At the end of the course students are able to

- Understand various different technologies related to database management system
- Understand when to use these technologies according to the requirements of particular applications
- Understand different alternative approaches proposed by extant database management systems for each of these technologies
- Understand the optimization issues related to particular implementation of these technologies in extant database management systems.

Readings and text books:

- R.T. Snodgrass. *Developing Time-Oriented Database Applications in SQL*, Morgan Kaufmann, 2000
- Jim Melton, Alan R. Simon. *SQL: 1999 - Understanding Relational Language Components*, Morgan Kaufmann, 2001
- Jim Melton. *Advanced SQL: 1999 - Understanding Object-Relational and Other Advanced Features*, Morgan Kaufmann, 2002
- Shashi Shekhar, Sanjay Chawla. *Spatial Databases: A Tour*, Prentice Hall, 2003.

Prerequisites:

- Knowledge of the basic principles of database management, in particular SQL

Table of contents:

- Active Databases
  Taxonomy of concepts. Applications of active databases: integrity maintenance, derived data, replication. Design of active databases: termination, confluence, determinism, modularisation.
- Temporal Databases
  Temporal data and applications. Time ontology. Conceptual modeling of temporal aspects. Manipulation of temporal data with standard SQL. New temporal extensions in SQL 2011.
- Object-Oriented and Object-Relational Databases
  Object-oriented model. Object persistance. ODMG standard: Object Definition Language and Object Query Language. .NET Language-Integrated Query: Linq.
  Object-relational model. Built-in constructed types. User-defined types. Typed tables. Type and table hierarchies. SQL standard and Oracle implementation.
- Spatial Databases
  Application domains of Geographical Information Systems (GIS), common GIS data types and analysis. Conceptual data models for spatial databases. Logical data models for spatial databases: rastor model (map algebra), vector model (OGIS/SQL1999). Physical data models for spatial databases: Clustering methods (space filling curves), storage methods (R-tree, Grid files).

Assessment breakdown:
75% written examination, 25% project evaluation

University: Université Libre de Bruxelles (ULB)
Department: École polytechnique de Bruxelles
Course ID: DBSA (INFO-H-417)
Course name: Database Systems Architecture
Name and email address of the instructors: Mahmoud Sakr (Mahmoud.Sakr@ulb.be)
Web page of the course: https://www.ulb.be/en/programme/info-h417
Semester: 1
Number of ECTS: 5

Course breakdown and hours:

- Lectures: 24h.
- Exercises: 12h.
- Projects: 24h.

Goals:

In contrast to a typical introductory course in database systems where one learns to design and query relational databases, the goal of this course is to get a fundamental insight into the implementation aspects of database systems. In particular, we take a look under the hood of relational database management systems, with a focus on query and transaction processing. By having an in-depth understanding of the query-optimisation-and-execution pipeline, one becomes more proficient in administering DBMSs, and hand-optimising SQL queries for fast execution.

Learning outcomes:

Upon successful completion of this course, the student:

- Understands the workflow by which a relational database management systems optimises and executes a query
- Is capable of hand-optimising SQL queries for faster execution
- Understands the I/O model of computation, and is capable of selecting and designing data structures and algorithms that are efficient in this model (both in the context of datababase systems, and in other contexts).
- Understands the manner in which relational database management systems provide support for transaction processing, concurrency control, and fault tolerance

Readings and text books:

- Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom. *Database Systems: The Complete Book*, Prentice Hall, second edition, 2008.
- Raghu Ramakrishnan, Johannes Gehrke. *Database Management Systems*. McGraw-Hill, third edition, 2002.

Prerequisites:

- Introductory course on relational databases, including SQL and relational algebra
- Course on algorithms and data structures
- Knowledge of the Java programming language

Table of contents:

- Query Processing
  With respect to query processing, we study the whole workflow of how a typical relational database management system optimises and executes SQL queries. This entails an in-depth study of:
  - translating the SQL query into a “logical query plan";
  - optimising the logical query plan;
  - how each logical operator can be algorithmically implemented on the physical (disk) level, and how secondary-memory index structures can be used to speed up these algorithms; and
  - the translation of the logical query plan into a physical query plan using cost-based plan estimation.
- Transaction Processing
  - Logging
  - Serializability
  - Concurrency control

Assessment breakdown:

70% written examination, 30% project evaluation

University: Université Libre de Bruxelles (ULB)
Department: École polytechnique de Bruxelles
Course ID: DW (INFO-H-419)
Course name: Data Warehousing
Name and email address of the instructors: Esteban Zimányi (ezimanyi@ulb.ac.be)
Web page of the course: http://cs.ulb.ac.be/public/teaching/infoh419
Semester: 1
Number of ECTS: 5

Course breakdown and hours:

- Lectures: 24h.
- Exercises: 24h.
- Projects: 12h.

Goals:

Relational and object-oriented databases are mainly suited for operational settings in which there are many small transactions querying and writing to the database. Consistency of the database (in the presence of potentially conflicting transactions) is of utmost importance. Much different is the situation in analytical processing where historical data is analyzed and aggregated in many different ways. Such queries differ significantly from the typical transactional queries in the relational model:

1. Typically analytical queries touch a larger part of the database and last longer than the transactional queries;
2. Analytical queries involve aggregations (min, max, avg, ...) over large subgroups of the data;
3. When analyzing data it is convenient to see it as multi-dimensional.

For these reasons, data to be analyzed is typically collected into a data warehouse with Online Analytical Processing support. Online here refers to the fact that the answers to the queries should not take too long to be computed. Collecting the data is often referred to as Extract-Transform-Load (ELT). The data in the data warehouse needs to be organized in a way to enable the analytical queries to be executed efficiently. For the relational model star and snowflake schemes are popular designs. Next to OLAP on top of a relational database (ROLAP), also native OLAP solutions based on multidimensional structures (MOLAP) exist. In order to further improve query answering efficiency, some query results can already be materialized in the database, and new indexing techniques have been developed.

The first and largest part of the course covers the traditional data warehousing techniques. The main concepts of multidimensional databases are illustrated using the SQL Server tools. The second part of the course consists of advanced topics such as data warehousing appliances, data stream processing, data mining, and spatial-temporal data warehousing. The coverage of these topics connects the data warehousing course with and serves as an introduction towards other related courses in the program. Several associated partners of the program contribute to the course in the form of invited lectures, case studies, and "proof of technology" sessions.

Learning outcomes:

At the end of the course students are able to

- Explain the difference between operational databases and data warehouses and the necessity of maintaining a dedicated data warehousing system
- Understand the principles of multidimensional modeling
- Design a formal conceptual multidimensional model based on an informal description of the available data and analysis needs
- Implement ETL-scripts for loading data from operational sources a the data warehouse
- Deploy data cubes and extract reports from the data warehouse
- Explain the main technological principles underlying data warehousing technology such as indexing and view materialization.

Readings and text books:

- Matteo Golfarelli, Stefano Rizzi. *Data Warehouse Design: Modern Principles and Methodologies.* McGraw-Hill, 2009
- Christian S. Jensen, Torben Bach Pedersen, Christian Thomsen. *Multidimensional Databases and Data Warehousing.* Morgan and Claypool Publishers, 2010
- Ralph Kimball, Margy Ross. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, third edition, Wiley, 2013.
- Esteban Zimányi, Alejandro Vaisman. *Data Warehouse Systems: Design and Implementation.* Springer,

2014

Prerequisites:

- A first course on database systems covering the relational model, SQL, entity-relationship modelling, constraints such as functional dependencies and referential integrity, primary keys, foreign keys.
- Data structures such as binary search trees, linked lists, multidimensional arrays.

Table of contents:

There is a mandatory project to be executed in three steps in groups of 3 students, using the tools learned during the practical sessions, being SQL Server, SSIS, SSAS, and SSRS. Below is the succinct summary of the theoretical part of the course:

- Foundations of multidimensional modelling
- Dimensional Fact Model
- Querying and reporting a multidimensional database with OLAP
- Methodological aspects for data warehouse development
- Populating a data warehouse: The ETL process
- Using the data warehouse: data mining and reporting

Assessment breakdown:

75% written examination, 25% project evaluation

University: Université Libre de Bruxelles (ULB)
Department: École polytechnique de Bruxelles
Course ID: BPM (INFO-H-420)
Course name: Management of Data Science and Business Workflows
Name and email address of the instructors: Dimitris Sacharidis (dimitris.sacharidis@ulb.be)
Web page of the course: https://www.ulb.be/en/programme/info-h420
Semester: 1
Number of ECTS: 5

Course breakdown and hours:

- Lectures: 24h.
- Exercises: 24h.
- Assignments: 12h.

Goals:

This course introduces basic concepts for managing workflows in data science applications and business processes. The first part of the course focuses on *business process management* and considers identification, modeling, analysis, simulation, redesign, and mining based on the Business Process Modeling and Notation (BPMN) workflow language. The second part focuses on *data science workflows* and discusses modeling, execution, and optimization, and also introduces various topics on *responsible data science*.

Learning outcomes:

At the end of the course students are able to:

- Explain the business process management cycle.
- Design a formal model of the business process based on an informal description.
- Identify opportunities for optimizing business processes.
- Describe data science workflows.
- Identify the costs associated with executing data science workflows.
- Optimize data science workflows.
- Identify concerns about data privacy and bias.
- Propose techniques to increase the explainability of data science workflows.

Readings and text books:

- Dumas, La Rosa, Mendling & Reijers. *Fundamentals of Business Process Management* (second edition), Springer 2018.

Prerequisites:

- Basic programming understanding.
- Basic set theory (notions such as set, set operations, sequence, multiset, function) and logics (mathematical notation and argumentation; basic proofs).
- Basic graph theory (notions such as graphs, reachability, transitivity).

Table of contents:

There are four assignments and a project to be realized in groups.

Below is a high-level overview of the theoretical part of the course:

- Business Process Management
  - Short overview of business processes and the need to manage them.
  - Describing business processes, modeling the control flow, data and resource perspectives.
  - Analysis of business processes, qualitatively and quantitatively.
  - Redesign of business processes.
  - Mining Process Logs.
- Data Science Workflows
  - Short overview of data science workflows.
  - Describing workflows in data science.
  - Analysis and optimization of data science workflows.
  - Data privacy.
  - Explainability of data science workflows.
  - Bias and fairness in data science workflows.

Assessment breakdown:
40% written examination, 60% assignment evaluation

University: Université Libre de Bruxelles (ULB)
Department: École polytechnique de Bruxelles
Course ID: DM (INFO-H-423)
Course name: Data Mining
Name and email address of the instructors: Mahmoud Sakr (Mahmoud.Sakr@ulb.be)
Web page of the course: http://cs.ulb.ac.be/public/teaching/infoh423
Semester: 1
Number of ECTS: 5

Course breakdown and hours:

- Lectures: 24h.
- Exercises: 24h.
- Projects: 12h.

Goals:

Data Mining aims at finding useful regularities in large structured and unstructured data sets. The goal of this course is to get a fundamental understanding of popular data mining techniques strengths and limitations, as well as their associated computational complexity issues. It will also identify industry branches which most benefit from Data Mining such as health care and e-commerce. The course will focus on business solutions and results by presenting case studies using real (public domain) data. The students will use recent Data Mining software.

Learning outcomes:

At the end of the course students are able to

- Establish the main characteristics and limitations of algorithms for addressing data mining tasks.
- Select, based on the description of a data mining problem, the most appropriate combination of algorithms to solve it.
- Develop and execute a data mining workflow on a real-life dataset to solve a data-driven analysis problem.
- Use recent data mining software for solving practical problems.
- Identify promising business applications of data mining.

Readings and text books:

- David J. Hand, Heikki Mannila, Padhraic Smyth. *Principles of Data Mining*. MIT Press, 2001.
- Delmater Rhonda, Hancock Monte. *Data Mining Explained*. Digital Press, 2001.
- Pang-Ning Tan, Michael Steinbach, Vipin Kumar. *Introduction to Data Mining*. Pearson Education (Addison Wesley), 2006.

Prerequisites:

- Programming experience
- Data structures
- Algorithms, complexity

Table of contents:

- Data mining
  - Data mining and knowledge discovery
  - Data mining functionalities
  - Data mining primitives, languages, and system architectures
- Data preprocessing
  - Data cleaning
  - Data transformation
  - Data reduction
  - Discretization and generating concept hierarchies
- Data mining algorithms
  - Motivation and terminology
  - Different algorithm types
- Classification and Clustering
  - Classification: SVM classifier
  - Clustering: K-means, Latent Semantic Analysis

- Mining Associations and Correlations
  - Item sets
  - Association rules
  - Generating item sets and rules efficiently
  - Correlation analysis
- Advanced techniques, data mining software, and applications
  - Text mining: extracting attributes (keywords), structural approaches (parsing, soft parsing).
  - Bayesian approach to classifying text
  - Web mining: classifying web pages, extracting knowledge from the web
  - Recommendation systems
  - Data mining software and applications

Assessment breakdown:
75% written examination, 25% project evaluation

University: Universitat Politècnica de Catalunya (UPC)
Department: Department of Service and Information System Engineering
Course ID: BDM
Course name: Big Data Management
Name and email address of the instructors: Alberto Abelló (aabello@essi.upc.edu)
Web page of the course: https://www.fib.upc.edu/en/studies/masters/master-innovation-and-research-informatics/curriculum/syllabus/BDM-MIRI
Semester: 2
Number of ECTS: 6

Course breakdown and hours:

- Lectures: 27 h.
- Laboratories: 27 h.
- Self-Study: 96 h.

Goals:
The main goal of this course is to analyze the technological and engineering needs of Big Data Management. The enabling technology for such a challenge is cloud services, which provide the elasticity needed to properly scale the infrastructure as the needs of the company grow. Thus, students will learn advanced data management techniques (i.e., NOSQL solutions) that also scale with the infrastructure. Being Big Data Management the evolution of Data Warehousing, such knowledge (see the corresponding subject in Data Science speciality for more details on its contents) is assumed in this course, which will specifically focus on the management of data Volume and Velocity.

On the one hand, to deal with high volumes of data, we will see how a distributed file system can scale to as many machines as necessary. Then, we will study different physical structures we can use to store our data in it. Such structures can be in the form of a file format at the operating system level, or at a higher level of abstraction. In the latter case, they take the form of either sets of key-value pairs, collections of semi-structured documents or column-wise stored tables. We will see that, independently of the kind of storage we choose, current highly parallelizable processing systems using functional programming principles (typically based on Map and Reduce functions), whose processing framework can rely on temporal files (like Hadoop MapReduce) or in-memory structures (like Spark).

On the other hand, to deal with high velocity of data, we need some low latency system which processes either streams or micro-batches. However, nowadays, data production is already beyond processing technologies capacity. More data is being generated than we can store or even process on the fly. Thus, we will recognize the need of (a) some techniques to select subsets of data (i.e., filter out or sample), (b) summarize them maximizing the valuable information retained, and (c) simplify our algorithms to reduce their computational complexity (i.e., doing one single pass over the data) and provide an approximate answer.

Finally, the complexity of a Big Data project (combining all the necessary tools in a collaborative ecosystem), which typically involves several people with different backgrounds, requires the definition of a high level architecture that abstracts technological difficulties and focuses on functionalities provided and interactions between modules. Therefore, we will also analyse different software architectures for Big Data.

This course participates in a joint project conducted during the second semester together with VBP, SDM and DEBD. In VBP, the students will come up with a business idea related to Big Data Management and Analytics, which will be evaluated from a business perspective. In DEBD, they should analyse the same idea from an ethical perspective. Finally, in BDM and SDM they will be introduce to specific data management techniques to deal with Volume and Velocity (BDM) and Variety (SDM). Therefore, as final outcome, a working prototype dealing with those challenges must be delivered meeting the business idea created.

Learning outcomes:
Upon successful completion of this course, the student is able to:

- Knowledge
  - Understand the main advanced methods of data management and design and implement non-relational database managers, with special emphasis on distributed systems.
  - Understand, design, explain and carry out parallel information processing in massively distributed systems.
  - Manage and process a continuous flow of data.
  - Design, implement and maintain system architectures that manage the data life cycle in analytical environments.
- Skills
  - Design a distributed database using NoSQL tools.

- Produce a functional program to process Big Data in a cloud environment.
- Manage and process a stream of data.
- Design the architecture of a Big Data management system.

Readings and text books:

- M. Tamer Özsu, Patrick Valduriez. *Principles of Distributed Database Systems*, Springer, 2011.
- Ling Liu, M. Tamer Özsu. *Encyclopedia of Database Systems*, Springer, 2009.
- Pramod J. Sadalage, Martin Fowler. *NoSQL Distilled*, Addison-Wesley, 2013.
- Mark Grover, Ted Malaska, Jonathan Seidman, Gwen Shapira. *Hadoop Application Architectures*, O'Rilley, 2015.
- Hasso Plattner, Alexander Zeier. *In-Memory Data Management*, Springer, 2011.
- Matei Zaharia. *An Architecture for Fast and General Data Processing on Large Clusters*, ACM Press, 2016.
- Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman. *Mining of Massive Datasets*, http://www.mmds.org/
- Charu C. Aggarwal (Ed.). *Data Streams*, Springer, 2007.

Prerequisites:

- Advanced Databases (ADB)
- Database Systems Architecture (DBSA)
- Data Warehousing (DW)

Table of contents:

I Fundamentals of distributed and in-memory databases

1 Introduction

- Cloud computing concepts
- Scalability

2 In-memory data management

- NUMA architectures

3 Distributed Databases

- Kinds of distributed databases
- Fragmentation
- Replication and synchronization (eventual consistency)
- Distributed query processing and Parallelism
- Bottlenecks of relational systems

II High volume management

4 Data management

- Physical structures (Key-values, Document-stores, and Column-stores)
- Distribution and placement

5 Data processing

- Functional programming

III High velocity management

6 Stream management

- Sliding window

7 Stream processing

- Sampling
- Filtering
- Sketching
- One-pass algorithms

IV Architectures

8 Software architectures

- Centralized and Distributed functional architectures of relational systems
- Lambda architecture

Assessment breakdown:

60% written examination, 40% project, +10% class participation

University: Universitat Politècnica de Catalunya
Department: Department of Service and Information System Engineering
Course ID: SDM
Course name: Semantic Data Management
Name and email address of the instructors: Oscar Romero (oromero@essi.upc.edu)
Web page of the course: https://www.fib.upc.edu/en/studies/masters/erasmus-mundus-master-big-data-management-and-analytics/curriculum/syllabus/SDM-BDMA
Semester: 2
Number of ECTS: 6

Course breakdown and hours:

- Lectures: 27 h.
- Laboratories: 27 h.
- Self-study: 96 h.

Goals:

Big Data is traditionally defined with the three V's: Volume, Velocity and Variety. Big Data has been traditionally associated with Volume (e.g., the Hadoop ecosystem) and recently Velocity has earned its momentum (especially, with the arrival of Stream processors such as Spark Streaming). However, even if Variety has been part of the Big Data definition, how to tackle Variety in real-world projects is yet not clear and there are no standarized solutions (such as Hadoop for Volume or Streaming for Velocity) for this challenge.

In this course the student will be introduced to advanced database technologies, modeling techniques and methods for tackling Variety for decision making. We will also explore the difficulties that arise when combining Variety with Volume and / or Velocity. The focus of this course is on the need to enrich the available data (typically owned by the organization) with external repositories (special attention will be paid to Open Data), in order to gain further insights into the organization business domain. There is a vast amount of examples of external data to be considered as relevant in the decision making processes of any company. For example, data coming from social networks such as Facebook or Twitter; data released by governmental bodies (such as town councils or governments); data coming from sensor networks (such as those in the city services within the Smart Cities paradigm); etc.

This is a new hot topic without a clear and well-established (mature enough) methodology. The student will learn about semantic-aware data management as the most promising solution for this problem. As such, it will be introduced to graph modeling, storage and processing. Special emphasis will be paid to semantic graph modeling (i.e., ontology languages such as RDF and OWL) and its specific storage and processing solutions.

This course participates in a joint project conducted during the second semester together. In VBP, the students will come up with a business idea related to Big Data Management and Analytics, which will be evaluated from a business perspective. In the three other courses the students have to implement a prototype meeting the business idea created. In BDM and SDM they will be introduce to specific data management techniques to deal with Volume and Velocity (BDM) and Variety (SDM). As final outcome, a working prototype must be delivered.

Learning outcomes:

The student will learn models, languages, methods and techniques to cope with Variety in the presence of Volume and Velocity. Specifically:

- Graph modeling, storage and processing,
- Semantic Models and Ontologies (RDF, RDFS and OWL),
- Logics-based foundations of ontology languages (Description Logics) and
- Specific storage and processing techniques for semantic graphs.

The students will learn how to apply the above mentioned foundations to automate the data management lifecycle in front of Variety. We will focus on semantic data governance protocols and the role of semantic metadata artifacts to assist the end-user.

Readings and text books:

- Serge Abiteboul, Ioana Manolescu, Philippe Rigaux, Marie-Christine Rousset, Pierre Senellart, *Web Data Management*, Cambridge University Press, 2011.
- Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, Peter F. Patel-Schneider, *The Description Logic Handbook*, Cambridge University Press, 2010.
- Sven Groppe, *Data Management and Query Processing in Semantic Databases*, Springer, 2011.

Prerequisites:

- Advanced Databases (ADB)
- Data Warehousing (DW)
- Database Systems Architecture (DBSA)

Table of contents:

- Introduction:
  - Variety and Variability. Definition. External (non-controlled) data sources. Semi-structured data models, non-structured data. Schema and data evolution.
  - The data management life-cycle. Challenges when incorporating external (to the organisation) semi-structured and non-structured data.
- Foundations:
  - Graph management: the graph data model, graph databases, graph processing.
  - Semantic-aware management: Ontology languages (RDF, RDFS, OWL). Logical foundations and Description Logics. Storage (triplestores). Processing (SPARQL).
- Applications:
  - Open Data. Linked Open Data.
  - The Load-First Model-Later paradigm. The Data Lake. Semantic annotations and the Semantic-Aware Data Lake paradigm. Semantic data governance.
  - Semantic-aware metadata artifacts to automate data integration, linkage and / or cross of data between heterogeneous data sources.

Assessment breakdown:
40% written examination, 50% exercises and laboratories, 10% project

University: Universitat Politècnica de Catalunya
Department: Department of Computer Science
Course ID: ML
Course name: Machine Learning
Name and email address of the instructors: Marta Arias (marias@cs.upc.edu)
Web page of the course: https://www.fib.upc.edu/en/estudis/masters/master-en-ciencia-de-dades/pla-destudis/assignatures/ML-MDS
Semester: 2
Number of ECTS: 6

Course breakdown and hours:

- Lectures: 27 h.
- Laboratories: 27 h.
- Self-study: 96 h.

Goals:

The aim of machine learning is the development of theories, techniques and algorithms to allow a computer system to modify its behavior in a given environment through inductive inference. The goal is to infer practical solutions to difficult problems –for which a direct approach is not feasible– based on observed data about a phenomenon or process. Machine learning is a meeting point of different disciplines: statistics, optimization and algorithmics, among others.

The course is divided into conceptual parts, corresponding to several kinds of fundamental tasks: supervised learning (classification and regression) and unsupervised learning (clustering, density estimation). Specific modelling techniques studied include artificial neural networks and support vector machines. An additional goal is getting acquainted with python and its powerful machine learning libraries.

Learning outcomes:

- Formulate the problem of (machine) learning from data, and know the different machine learning tasks, goals and tools.
- Organize the workflow for solving a machine learning problem, analyzing the possible options and choosing the most appropriate to the problem at hand.
- Ability to decide, defend and criticize a solution to a machine learning problem, arguing the strengths and weaknesses of the approach. Additionally, ability to compare, judge and interpret a set of results after making a hypothesis about a machine learning problem.
- To be able to solve concrete machine learning problems with available open-source software.

Readings and text books:

- C.M. Bishop, *Pattern recognition and machine learning*, Springer, 2006.
- V.S. Cherkassky, F. Mulier, *Learning from data: concepts, theory, and methods*, John Wiley, 2007.
- E. Alpaydin, *Introduction to machine learning*, The MIT Press, 2014.
- K.P. Murphy, *Machine learning: a probabilistic perspective*, MIT Press, 2012.

Prerequisites:

- Elementary notions of probability and statistics.
- Elementary linear algebra and real analysis.
- Good programming skills in a high-level language.

Table of contents:

- Introduction to Machine Learning.
  General information and basic concepts. Overview to the problems tackled by machine learning techniques. Supervised learning (classification and regression), unsupervised learning (clustering and density estimation) and semi-supervised learning (reinforcement and transductive). Examples.
- Supervised machine learning theory.
  The supervised Machine Learning problem setup. Classification and regression problems. Bias-variance tradeoff. Regularization. Overfitting and underfitting. Model selection and resampling methods.
- Linear methods for regression.
  Error functions for regression. Least squares: analytical and iterative methods. Regularized least squares. The Delta rule. Examples.
- Linear methods for classification.
  Error functions for classification. The perceptron algorithm. Novikoff's theorem. Separations with maxi-

mum margin. Generative learning algorithms and Gaussian discriminant analysis. Naive Bayes. Logistic regression. Multinomial regression.

- Artificial neural networks.
  Artificial neural networks: multilayer perceptron and a peak into deep learning. Application to classification and to regression problems.
- Kernel functions and support vector machines.
  Definition and properties of Kernel functions. Support vector machines for classification and regression problems.
- Unsupervised machine learning.
  Unsupervised machine learning techniques. Clustering algorithms: EM algorithm and k-means algorithm.
- Ensemble methods.
  Bagging and boosting methods, with an emphasis on Random Forests.

40% final exam, 40% practical work and 20% mid-term exam.

University: Universitat Politècnica de Catalunya (UPC)
Department: Department of Management
Course ID: VBP
Course name: Viability of Business Projects
Name and email address of the instructors: Marc Eguiguren (marc.eguiguren@upc.edu)
Web page of the course: https://www.fib.upc.edu/en/studies/masters/master-innovation-and-research-informatics/curriculum/syllabus/VBP-MIRI
Semester: 3
Number of ECTS: 6

Course breakdown and hours:

- Lectures: 36 h.
- Projects in the classroom: 18 h.
- Projects: 56 h.
- Self-Study: 40 h.

Goals:

University graduates can find themselves in the situation of having to analyse or take on the project of starting their own business. This is especially true in the case of computer scientists in any field related to Big Data Management (BDM) or more generally, in the world of services. There are moments in one's professional career at which one must be able to assess or judge the suitability of business ventures undertaken or promoted by third parties, or understand the possibilities of success of dealing with a Big Data based service. It is for this reason that this subject focuses on providing students with an understanding of the main techniques used in analysing the viability of new business ventures: business start-up or the implementation of new projects in the world of services based on BDM. This project-oriented, eminently practical subject is aimed at each student's being able to draft as realistic a business plan as possible.

This course participates in a joint project conducted during the second semester together with VBP, BDM and CC. In VBP, the students will come up with a business idea related to Big Data Management and Analytics, which will be evaluated from a business perspective. In the three other courses the students have to implement a prototype meeting the business idea created. In CC they will be introduced to the main concepts behind large-scale distributed computing based on a service-based model and will have to choose the right infrastructure for their prototype. In BDM and SDM they will be introduce to specific data management techniques to deal with Volume and Velocity (BDM) and Variety (SDM). As final outcome, a working prototype must be delivered.

Learning outcomes:

Upon successful completion of this course, the student is able to:

- Being able to analyze the external situation to determine business innovative ideas in the field of BDM
- Around an innovative BDM project, being able to build a reasonable and ethically solid business plan
- Building a solid and convincing speech about a business idea and a business plan
- Training the students to build a P&L forecast and a forecasted treasury plan for a starting company
- Understanding and being able to apply the different instruments to finance the company, both debt instruments or private equity and venture capital sources
- Understand and appreciate the role of the entrepreneur in modern society

Readings and text books:

- Rhonda Abrams, Eugène Kleiner. *The Successful Business Plan.* The Planning Shop, 2003.
- Rob Cosgrove. *Online Backup Guide for Service Providers*, Cosgrove, Rob, 2010.
- Peter Drucker. *Innovation and Entrepreneurs.* Butterworth-Heinemann, Classic Drucker Collection edition, 2007.
- Robert D. Hisrich, Michael P. Peter, Dean A. Shepherd. *Entrepreneurship.* Mc Graw Hill, 6^{th} Ed., 2005.
- Mike McKeever. *How to Write a Business Plan.* Nolo, 2010.
- Lawrence W. Tuller. *Finance for Non-Financial Managers and Small Business Owners.* Adams Business, 2008.

Other Literature:

- M. Eguiguren, E. Barroso. *Empresa 3.0: políticas y valores corporativos en una cultura empresarial sostenible.* Pirámide , 2011.
- M. Eguiguren, E. Barroso. *Por qué fracasan las organizaciones: de los errores también se aprende.* Pirámide, 2013.

Prerequisites:
Having some previous knowledge or experience in business administration is an additional asset.

Table of contents: This course focuses on developing a BDM service oriented business plan. So that, the students are expected to reuse and consolidate any previous knowledge on databases, software engineering and BDM obtained in previous courses to develop a comprehensible, sustainable and profitable business.

The course is structured in 14 well-defined stages:

- Introduction to the course and key aspects of a business idea
- The entrepreneur's role in society, characteristics and profile
- Innovation and benchmarking Axis 1) Identification of long-term market megatrends
- Innovation and benchmarking axis 2) Big Data evolution as a source of ideas. Technology applied to industry.
- Axis of innovation and benchmarking 3) ethical business models as a source of innovation and ideas
- From the idea to the company. Contents of the business plan. Market research.
- Competitive advantages. SWOT Analysis
- Marketing plan: strategic marketing for a BDM service company, distribution and product
- Marketing plan: price and promotion strategies
- The human team in a small innovative company
- Different kind of societies. Fiscal basics for entrepreneurs
- Need of resources. Building the balance sheet at the beginning of the company
- Building a forecasted P&L for the first two years. Cash-Flow
- Revising the initial balance sheet and building the forecasted balance sheet for year one
- Treasury plan, Identifying long and short term financial needs
- Conventional long and short term financial instruments
- Private equity: founders, fools, friends & family, venture capital. Their limitations. Cautions to be taken and how they work.
- Presenting the plan to possible simulated or real investors

The business plan is expected to be partially developed in internal activities (under supervision of the teacher), and in external activities, always as teamwork (with no supervision).

Assessment breakdown:
The assessment is based on student presentations and the defence of the business plan before a jury comprising course faculty members and - optionally - another member of the teaching staff or guest professionals such as business angels, investors and successful IT enterpreneurs.

Throughout the course there will be four evaluative milestones:

- presentation of the innovative business model,
- presentation of the marketing plan,
- presentation of the business plan as a whole, that will include an evaluation about ethics and sustainability of the project together with SEAIT,
- analysis of the financial plan and the proposal to investors.

The presentation simulates a professional setting. Accordingly, the following aspects will also be assessed: dress, formal, well-structured communication, etc.

In order to be able to publicly defend the business plan, students must have attended at least 70% of the classes and teams must have delivered on time the activities that have been planned during the course. The plan is the result of teamwork, which will be reflected in the grade given to the group as a whole. Each member of the group will be responsible for part of the project and will be graded individually on his or her contribution.

This approach is designed to foster teamwork, in which members share responsibility for attaining a common objective.

University: Universitat Politècnica de Catalunya (UPC)
Department: Department of Service and Information System Engineering
Course ID: BDS
Course name: Big Data Seminar
Name and email address of the instructors: Oscar Romero (oromero@essi.upc.edu)
Web page of the course: To be created https://www.fib.upc.edu/en/studies/masters/erasmus-mundus-master-big-data-management-and-analytics/curriculum/syllabus/BDS-BDMA
Semester: 2
Number of ECTS: 2

Course breakdown and hours:

- Lectures: 21 h.
- Autonomous work: 129 h.

Goals:
The students will be introduced to recent trends in Big Data. Seminars will be lectured by guest speakers, who will present business cases, research topics, internships and master's thesis subjects. Also, the second year specialisations will be presented and discussed with the students within the seminars umbrella. Students will also perform a state-of-the art research in one topic, which will be presented and jointly evaluated by all partners in the mandatory eBISS summer school.

Learning outcomes:
Upon successful completion of this course, the student is able to:

- Read and understand scientific papers
- Develop critical thinking when assessing scientific papers
- Write and explain a state-of-the-art in a rigorous manner
- Elaborate on recent trends in Big Data

Readings and text books:

- Gordana Dodig-Crnkovic, *Theory of Science*, online resource: http://www.idt.mdh.se/kurser/ct3340/archives/ht04/theory_of_science_compendiu.pdf
- Jennifer Widom, *Tips for Writing Technical Papers*, online resource: https://cs.stanford.edu/people/widom/paper-writing.html
- Robert Siegel, *Reading Scientific Papers*, online resource: https://web.stanford.edu/~siegelr/readingsci.htm

Prerequisites:

- No prerequisites

Table of contents:
The seminars content will vary from course to course as they will focus on current hot topics in Big Data.

Assessment breakdown:
50% Written report on a chosen state-of-the-art,
50% Poster presentation.
Note: Attendance to the eBISS summer school is mandatory. To be evaluated (see the formula above) students must guarantee >75% overall attendance to the summer school events and the semester seminars.

University: Universitat Politècnica de Catalunya (UPC)
Department: Department of Service and Information System Engineering
Course ID: DEBD
Course name: Debates on Ethics of Big Data
Name and email address of the instructors: Alberto Abelló (aabello@essi.upc.edu)
Web page of the course: https://www.fib.upc.edu/en/studies/masters/erasmus-mundus-master-big-data-management-and-analytics/curriculum/syllabus/DEBD-BDMA
Semester: 2
Number of ECTS: 2

Course breakdown and hours:

- Lectures: 18 h.
- Autonomous work: 32 h.

Goals:
In this course we debate the impact on society of new advances in Big Data. We focus on ethics and the impact of such approaches on society. This course fosters the social competences of students, by building on their acquired oral communication skills to debate about concrete problems involving ethical issues in Big Data. The aim is to start developing their critical attitude and reflection. A written summary of their position is meant to train their writing skills.

During the course sessions the debates discussing innovative and visionary ideas on Big Data will take place. You must read the available material before the debate. Then, during the debate you will be assign to a group: either to defend an idea, or go against it. You may also be asked to moderate the debate. Then, the debate takes place and afterwards, each group needs to write down a report with their conclusions.

Learning outcomes:
Upon successful completion of this course, the student will develop:

- Ability to study and analyze problems in a critical mood
- Ability to critically read texts
- Develop critical reasoning with special focus on ethics and social impact
- Develop soft skills to defend - criticize a predetermined position in public
- Improve the writing skills

Readings and text books:

- Rudi Volti, *Society and technological change*, Worth, 2009.
- Richard T. De George, *The Ethics of information technology and business*, Blackwell, 2003.
- David Elliott, *Energy, society, and environment: technology for a sustainable future*, Routledge, 2003.
- Kord Davis, *Ethics of big data*, O'Reilly, 2013.

Prerequisites:

- No prerequisites

Table of contents:
Debates on ethics and social impact of Big Data (some examples):

- Ethics codes
- Right to privacy
- Content piracy
- Social responsibility of IT companies
- Artificial Intelligence and their limits

Assessment breakdown:
80% Debate evaluations, 20% Ethical analysis of a data-based business idea

University: Universitá degli Studi di Padova (UniPd)
Department: Dipartimento di Matematica
Course ID: StatLearn
Course name: Statistical Learning
Name and email address of the instructors: ALBERTO ROVERATO (alberto.roverato@unipd.it)
Web page of the course: https://en.didattica.unipd.it/off/2021/LM/SC/SC2377/001PD/SCP7079227/N0
Semester: 1
Number of ECTS: 6

Lectures, Exercises: 48h.

Goals:
Become familiar with statistical thinking; gain adequate proficiency in the development and use of standard statistical inference tools; be able to analyse datasets using a modern programming language such as R

Learning outcomes:
At the end of the course students should show knowledge of the key concepts, skills in the analysis of data and competency in applications.

Readings and text books:

- Hastie, T., Tibshirani, R., and Friedman, J., The Elements of Statistical Learning: Data Mining, Inference, and Prediction. –: Springer, 2001. https://web.stanford.edu/~hastie/ElemStatLearn/
- Lavine, M., Introduction to Statistical Thought. –: None, 2013. http://people.math.umass.edu/~lavine/Book/book.html
- Gareth, James, An Introduction to Statistical Learning. –: Springer, 2013. https://www-bcf.usc.edu/~gareth/ISL/

Additional notes on the applications can be found in:

- Nolan, D.A. & Speed, T. (2000). Stat Labs: Mathematical Statistics Through Applications. Springer.
- Torgo, L. (2011). Data Mining with R: Learning with Case Studies. Chapman & Hall/CRC.

Methods for specific fields of applications can be found in the following books:

- Campbell, R.C. (1989). Statistics for Biologists (3rd ed.). Cambridge University Press.
- Devore, J.L. (2000). Probability and Statistics for Engineering and the Sciences (5th ed.). Duxbury Press, Pacific Grove, CA.
- Agresti, A. & Finlay. B. (2007). Statistical Methods for the Social Sciences (4th ed.). Prentice Hall

Prerequisites:

- Basic probability theory; multivariable calculus; linear algebra; basic computing skills.

Table of contents:

- Data: summary statistics, displaying distributions; exploring relationships
- Estimation: point estimation; the sampling distribution of an estimator; accuracy of estimation; interval estimation
- Hypothesis testing
- Likelihood: the likelihood, likelihood for several parameters
- Estimation: maximum likelihood estimation; properties of maximum likelihood estimates

Assessment breakdown:
Written examination, and project work

University: Universitá degli Studi di Padova (UniPd)
Department: Dipartimento di Matematica
Course ID: DeepLearn
Course name: Deep Learning and Human Data Analytics
Name and email address of the instructors: MICHELE ROSSI (michele.rossi@unipd.it)
Web page of the course: https://en.didattica.unipd.it/off/2021/LM/SC/SC2377/000ZZ/SCP7079397/N0 (This course is shown under a different name, but is renamed as intended in this syllabus for BDMA)
Semester: 1
Number of ECTS: 6

Course breakdown and hours:

- Lectures: 36h.
- Exercises and Project: 12h.

Goals:
Introduce advanced machine- and deep-learning techniques, and apply them to real-world problem within the human-data domain

Learning outcomes:
At the end of the course students will acquire skills related to:

1. the main clustering algorithms for vector data, their pros and cons, their evaluation metrics
2. the main unsupervised learning techniques for unsupervised vector quantization, their performance, advantages and their use within real problems involving biosignals
3. the main modeling techniques for multivariate time series and their use within selected applications in the human data domain
4. the principles and main algorithms for supervised learning through neural networks (feed forward, convolutional and recurrent), their programming in Python, and their use to solve real world problems
5. the main application domains for the techniques at the previous points 1, 2, 3 and 4 and how they can be exploited to tackle relevant problems within the "human data" domain
6. comprehending, selecting and knowing how to use the techniques at the previous points 1, 2, 3 and 4
7. solving a real world problem involving human data analysis, through summarizing its solution through a professional written report, presenting the work done via a conference-style talk, also showcasing the software written for the project
8. implementing and use the techniques at points 1, 2, 3 and 4 in Python

Readings and text books:

- Bishop, Christopher M., Pattern recognition and machine learning. Christopher M. Bishop. New York: Springer, 2006.
- Bengio, Yoshua; Courville, Aaron; Goodfellow, Ian, Deep Learning. Cambridge: MIT Press, 2016.
- Watt, Jeremy; Borhani, Reza; K. Katsaggelos, Aggelos, Machine Learning Refined: Foundations, Algorithms, and Applications. New York: Cambridge University Press, 2020
- Technical reports, scientific papers

Prerequisites: Prior knowledge on Calculus and Linear Algebra, and some basic computer programming is useful, preferably in Python. Prospective students will benefit from prior attendance of a basic Machine-Learning course.

Table of contents:

- Vector quantization (VQ):K-means, soft K-means, Expectation Maximization (EM), X-means, DBSCAN (density based clustering)
- Unsupervised VQ algorithms: Self-Organizing Maps (SOM), Gas Neural Networks (GNG): theory and algorithms
- Application to quasi-periodic biometric signals (ECG): Signal pre-processing, normalization, segmentation, efficient representation of ECG, final system design and numerical results
- Deep Neural Networks: gradient descent, Feed Forward Neural Networks: models, training, back-propagation, Convolutional Neural Networks (CNN), Autoencoder architectures for unsupervised feature extraction, denoising autoencoders Recurrent neural networks (RNN) for the analysis of temporal sequences, Residual convolutional neural networks (ResNets), Attention mechanisms for CNN (images)
- Use of Keras and TensorFlow frameworks in Python for the implementation and training of neural network structures.

- Training of neural network structured on real datasets

Assessment breakdown:
The student examinations will be carried out through a project. The grade is determined on the basis of a written report and the oral discussion of the project.

University: Universitá degli Studi di Padova (UniPd)
Department: Dipartimento di Matematica
Course ID: TimeSeries
Course name: Time-Series Analysis for Business Economic and Financial Data
Name and email address of the instructors: MICHELE ROSSI (michele.rossi@unipd.it)
Web page of the course: https://en.didattica.unipd.it/off/2021/LM/SC/SC2377/000ZZ/SCP7079397/N0 (This course is shown under a different name, but is renamed as intended in this syllabus for BDMA)
Semester: 1
Number of ECTS: 6

Course breakdown and hours:

- Lectures: 36h.
- Exercises and Project: 12h.

Goals:
Introduce advanced machine- and deep-learning techniques, and apply them to real-world problem within the human-data domain

Learning outcomes:
At the end of the course students will acquire skills related to:

1. the main clustering algorithms for vector data, their pros and cons, their evaluation metrics
2. the main unsupervised learning techniques for unsupervised vector quantization, their performance, advantages and their use within real problems involving biosignals
3. the main modeling techniques for multivariate time series and their use within selected applications in the human data domain
4. the principles and main algorithms for supervised learning through neural networks (feed forward, convolutional and recurrent), their programming in Python, and their use to solve real world problems
5. the main application domains for the techniques at the previous points 1, 2, 3 and 4 and how they can be exploited to tackle relevant problems within the "human data" domain
6. comprehending, selecting and knowing how to use the techniques at the previous points 1, 2, 3 and 4
7. solving a real world problem involving human data analysis, through summarizing its solution through a professional written report, presenting the work done via a conference-style talk, also showcasing the software written for the project
8. implementing and use the techniques at points 1, 2, 3 and 4 in Python

Readings and text books:

- Bishop, Christopher M., Pattern recognition and machine learning. Christopher M. Bishop. New York: Springer, 2006.
- Bengio, Yoshua; Courville, Aaron; Goodfellow, Ian, Deep Learning. Cambridge: MIT Press, 2016.
- Watt, Jeremy; Borhani, Reza; K. Katsaggelos, Aggelos, Machine Learning Refined: Foundations, Algorithms, and Applications. New York: Cambridge University Press, 2020
- Technical reports, scientific papers

Prerequisites: Prior knowledge on Calculus and Linear Algebra, and some basic computer programming is useful, preferably in Python. Prospective students will benefit from prior attendance of a basic Machine-Learning course.

Table of contents:

- Vector quantization (VQ):K-means, soft K-means, Expectation Maximization (EM), X-means, DBSCAN (density based clustering)
- Unsupervised VQ algorithms: Self-Organizing Maps (SOM), Gas Neural Networks (GNG): theory and algorithms
- Application to quasi-periodic biometric signals (ECG): Signal pre-processing, normalization, segmentation, efficient representation of ECG, final system design and numerical results
- Deep Neural Networks: gradient descent, Feed Forward Neural Networks: models, training, back-propagation, Convolutional Neural Networks (CNN), Autoencoder architectures for unsupervised feature extraction, denoising autoencoders Recurrent neural networks (RNN) for the analysis of temporal sequences, Residual convolutional neural networks (ResNets), Attention mechanisms for CNN (images)
- Use of Keras and TensorFlow frameworks in Python for the implementation and training of neural network structures.

- Training of neural network structured on real datasets

Assessment breakdown:
The student examinations will be carried out through a project. The grade is determined on the basis of a written report and the oral discussion of the project.

University: Eindhoven University of Technology (TU/e)
Department: Department of Mathematics and Computer Science
Course ID: 2AMI10
Course name: Foundations of Process Mining
Name and email address of the instructors: prof.dr.ir. Boudewijn van Dongen (b.f.v.dongen@tue.nl)
Web page of the course:
https://tue.osiris-student.nl/#/onderwijscatalogus/extern/cursus?cursuscode=2AMI10&collegejaar=2022&taal=en
Semester: 3 (Q1)
Number of ECTS: 5

Course breakdown and hours:

- Lectures: 32 h.
- Instructions: 32 h.
- Self-Study: 76 h.

Goals:
Data science is the profession of the future, because organizations that are unable to use (big) data in a smart way will not survive. It is not sufficient to focus on data storage and data analysis. The data scientist also needs to relate data to process analysis. Process mining bridges the gap between traditional model-based process analysis (e.g., simulation and other business process management techniques) and data-centric analysis techniques such as machine learning and data mining. Process mining seeks the confrontation between event data (i.e., observed behavior) and process models (hand-made or discovered automatically). This technology has become available only recently, but it can be applied to any type of operational processes (organizations and systems). Example applications include: analyzing treatment processes in hospitals, improving customer service processes in a multinational, understanding the browsing behavior of customers using a booking site, analyzing failures of a baggage handling system, and improving the user interface of an X-ray machine. All of these applications have in common that dynamic behavior needs to be related to process models. Hence, we refer to this as "data science in action".

The course covers the fundamentals of process mining. We start with data pre-processing and we show how through a visual inspection of event data and filtering, data can be pre-processed. In a second step, we consider the problem of process discovery, i.e. how to obtain a process model from an event log automatically. We again focus on the basics, but also show state-of-the-art in this area. To assess the quality of the discovered models, we show the fundamentals of conformance checking, i.e. the comparison of event logs and process models. We discuss the four main quality dimensions: fitness, precision, generalization and simplicity.

Finally, we show how the event log can be combined with the model to enrich the model with other perspectives, such as performance, decision points and resource information. The course uses many examples using real-life event logs to illustrate the concepts and algorithms. After taking this course, one is able to run process mining projects and have a good understanding of the data science field.

Learning outcomes:
After taking this course students should:

- have a good understanding of process mining,
- understand the role of data science in today's society,
- be able to relate process mining techniques to other analysis techniques such as simulation, business intelligence, data mining, machine learning, and verification,
- be able to apply basic process discovery techniques to learn a process model from an event log,
- be able to apply basic conformance checking techniques to compare event logs and process models,
- be able to extend a process model with information extracted from the event log (e.g., show bottlenecks),
- have a good understanding of the data needed to start a process mining project,
- be able to clean and reshape data to match the requirements for a process mining project

Readings and text books:

- W. van der Aalst. Process Mining: Data Science in Action. Springer-Verlag, Berlin, 2016 (https://link.springer.com/book/10.1007/978-3-662-49851-4%C2%A0). ISBN: 978-3-662-49850-7
- Carmona, J., van Dongen, B., Solti, A., Weidlich, M. Conformance Checking (https://link.springer.com/book/10.1007%2F978-3-319-99414-7) ISBN: 978-3-319-99414-7

Prerequisites:

- In Block GS1 you may not be part of one of the following target groups
  - Due to overlap with 2IMI35 you cant enrol for 2AMI10

Assessment breakdown:
100% Written final exam (minimum grade 5.0)

University: Eindhoven University of Technology (TU/e)
Department: Department of Mathematics and Computer Science
Course ID: 2AMS10
Course name: Longitudinal Data Analysis
Name and email address of the instructors: prof.dr. Edwin van den Heuvel (e.r.v.d.heuvel@tue.nl)
Web page of the course:
https://tue.osiris-student.nl/#/onderwijscatalogus/extern/cursus?cursuscode=2AMS10&collegejaar=2022&taal=en
Semester: 3 (Q1)
Number of ECTS: 5

Course breakdown and hours:

- Lectures: 32 h.
- Instructions: 32 h.
- Self-Study: 76 h.

Goals:
Longitudinal data is data collected over time (repeated measures) on a set of units to be able to study changes on these units over time. It is common in a wide range of disciplines, like economics, life sciences, and industry. Longitudinal data typically shows differences between units, differences over time, and correlations between repeated observations over time. The course LDA focuses on statistical techniques for analyzing longitudinal data where only a limited number of repeated observations on units are being collected (making it quite different from time series analysis where a lot of repeated data on each unit or one unit is collected). The statistical techniques are focused on understanding and estimating the different sources of variability in longitudinal data (unit-to-unit; within-unit; and time-changes). Topics that are included are:

- Hypothesis testing (e.g. differences between groups of units, between different time points, and multiple testing)
- Estimation and confidence intervals (e.g. correlations, variance components, and time changes)
- Statistical modelling (e.g. analysis of variance and mixed effects models)
- Diagnostics and goodness-of-fit (e.g. outlier detection, homogeneity of variances, normality testing, and information criteria)

The topics will be discussed in the context of real longitudinal data with real research questions to be answered. The SAS software (university edition) will be used to be able to analyze the data sets.

Learning outcomes:
After taking this course students should be able to:

- Understand and explain the different complexities that exist within longitudinal data (with respect to the application domain)
- Explain, discuss, criticize, and communicate results from (any) longitudinal data analysis
- Understand the role and limitations of statistical analysis methods for analyzing longitudinal data (in particular in the health domain)
- Analyze and model longitudinal data with SAS software
- Evaluate goodness-of-fit for the selected statistical model on longitudinal data

Readings and text books:

- Tutorial papers on longitudinal data analysis
- Sheskin D, Handbook of parametric and nonparametric statistical procedures, Chapman & Hall, 5th edition, 2011
- Fitzmaurice GM, Laird NM, Ware JH, Applied Longitudinal Analysis, Wiley, 2nd edition, 2011

Prerequisites:

- In Block GS1 you may not be part of one of the following target groups
  - no 2DMT00 or 2MMS70 enrollment
- Completed none of the course modules listed below
  - Applied statistics (2DMT00)
  - Statistical analysis methods (2MMS70)

Assessment breakdown:
70% Written final exam (minimum grade 5.0)
30% Assignment

University: Eindhoven University of Technology (TU/e)
Department: Department of Mathematics and Computer Science
Course ID: 2AMR10
Course name: Responsible Data Challenge
Name and email address of the instructors: dr. Dirk Fahland (d.fahland@tue.nl)
Web page of the course:
https://tue.osiris-student.nl/#/onderwijscatalogus/extern/cursus?cursuscode=2AMR10&collegejaar=2022&taal=en
Semester: 3 (Q1)
Number of ECTS: 5

Course breakdown and hours:

- Lectures: 32 h.
- Instructions: 32 h.
- Self-Study: 76 h.

Goals:

Learning outcomes:
After taking this course students should be able to:

-

Readings and text books:

-

Prerequisites:

- You must meet the following requirements
  - Enrolled for a degree programme of faculty Mathematics and Computer Science
  - Enrolled for one of the following degree programmes
    - Data Science in Engineering (CSE)

Assessment breakdown:
100% Assignment

University: Eindhoven University of Technology (TU/e)
Department: Department of Mathematics and Computer Science
Course ID: 2AMI20
Course name: Advanced Process Mining
Name and email address of the instructors: dr. Dirk Fahland (d.fahland@tue.nl)
Web page of the course: https://tue.osiris-student.nl/#/onderwijscatalogus/extern/cursus?cursuscode=2AMI20&collegejaar=2021&taal=en
Semester: 3 (Q2)
Number of ECTS: 5

Course breakdown and hours:

- Lectures: 32 h.
- Instructions: 32 h.
- Self-Study: 76 h.

Goals:
Understanding and predicting behavior of people and machines in a shared setting (task, project, factory, organization) is central to Data Science and Artificial Intelligence. Actions of people and machines can be recorded as discrete events in event sequences (logs), event databases (tables, graphs), and real-time event streams. Learning behavioral models of discrete event data of human behavior is challenging. Only those events which are causally related may be analyzed together. Further, the analysis results must be fully explainable and interpretable by humans, to evaluate, understand, communicate and improve the model - to let users take correct decisions in concrete situations.

This advanced course on process mining teaches students the fundamental concepts and theoretical foundations of process mining along a complete process mining methodology, and exposes students to real-life data sets to understand challenges related to process discovery, conformance checking, and model extension. The course material is based on recent research artciels in the field and the course teaches students how to read and understand research literature. The course is organized in two parts:

The common part (weeks 1-3) covers

- foundational understanding of event data, constructing new event logs from raw event data, and how to obtain obtain insights into complex event data through pre-processing and data visualization
- foundational event data abstraction techniques used in most model learning methods for descripte and predictive models, how different forms of abstraction impact model quality (accuracy, generalization, and understandability) and how these are used in selected model learning algorithms
- critically evaluating behavioral models (descriptive and predictive) regarding their behavioral accuracy and their explainability (generalizability and understandability) and how to use detailed diagnostic information at the level of individual events for model evolution, repair, and for drawing actionable conclusions

In the second part (weeks 4-7), students follow one of the following specialization tracks

- systematically designing model learning algorithms with quality guarantess; improving machine-learning prediction models through feature engineering from event data based on qualitative model evaluation; using event graphs for in-database process mining (recording, querying, and modeling) of entire system-level dynamics across multiple processes
- process mining on (real-time) event streams and stream data mining, and
- process mining on unstructured event data for knowledge-driven processes.

All concepts will be discussed and illustrated on concrete cases and event datasets from a variety of domains, including hospitals, high-tech systems, logistics systems, insurance companies, governments, etc.

The course is taught as a flipped classroom course with a group-assignment:

- Each topic has a concrete, hands-on reading assignment prior to class; you can use the social reading platform www.perusall.com where you can annotate text passages and discuss with fellow students and lecturers the parts you find difficult to understand.
- We provide for each topic practical exercises: paper-based exercises on the level of the final exam, and tool-based exercises to try out and understand the techniques.
- During the online class (video meetings), we specifically address your questions about concepts and ideas you found difficult to understand from reading. We solve example assignments and explain the concepts interactively.
- In the group assignment, you analyze a real-life dataset using a structured data science analysis methodology and the techniques taught in the course to discover, evaluate, and improve explainable models and to draw actionable conclusion.

The course replaces 2IMI20 Advanced Process Mining. In comparison, it focuses more on advanced topics

and ongoing research. Students must have passed Introduction to Process Mining (2IMI35) or Foundations of Process Mining (2AMI10) to participate.

Learning outcomes:
After taking this course students should be able to:

- have a detailed understanding of the entire process mining spectrum and the methodology for process mining analysis
- ... "good engineer"
- can derive and pre-process event logs from raw data and have understand and can work with a specialized form of event data such as unstructured event data in event logs, event graphs, or real-time event streams
- have a detailed understanding and be able to explain various concepts for learning models from event data and their specific properties and limitations in relation to the event data properties, specifically recognizing temporal patterns and constraints from event sequences, inferring behavioral and causal relations from aggregations of event data, handling concept drift, outlier detection and stream-mining concepts in comparison to static techniques.
- have a detailed understanding of quality criteria for explainable models and how explainable process models complement general purpose machine learning models, and can evaluate models regarding behavioral accuracy, generalizability, and understandability for event logs and as well as unstructured event data, multi-dimensional event data, or event streams
- independently execute a process mining analysis and critically compare and apply various process mining techniques to discover, evaluate, and improve explainable behavioral models and to draw actionable conclusion

Readings and text books:

- Papers, slides, event logs, and exercises provided via LMS.
- The textbook W. van der Aalst. Process Mining: Data Science in Action. Springer-Verlag, Berlin, 2016 (http://springer.com/978-3-662-49850-7). Available via: https://link.springer.com/book/10.1007/978-3-662-49851-4
- The book "Conformance Checking" by J. Carmona, B.F. van Dongen, M. Weidlich and A. Solti, Springer-Verlag, Cham, 2018. Available via: https://link.springer.com/book/10.1007/978-3-319-99414-7

Prerequisites:

- You must meet the following requirements
  - Enrolled for one of the following degree programmes
    - Data Science and Artificial Intelligence
    - Data Science in Engineering (CSE)
    - Operations Management and Logistics
  - In Block GS2 you may not be part of one of the following target groups
    - Due to overlap with 2IMI20 you cant enrol for 2AMI20

Assessment breakdown:
60% Written final exam (minimum grade 5.0)
40% Assignment

University: Eindhoven University of Technology (TU/e)
Department: Department of Mathematics and Computer Science
Course ID: 2IMP40
Course name: Applications of Data Science for Software Engineering
Name and email address of the instructors: prof.dr. Alexander Serebrenik (a.serebrenik@tue.nl)
Web page of the course:
https://tue.osiris-student.nl/#/onderwijscatalogus/extern/cursus?cursuscode=2IMP40&collegejaar=2022&taal=en
Semester: 3 (Q2)
Number of ECTS: 5

Course breakdown and hours:

- Lectures: 32 h.
- Instructions: 32 h.
- Self-Study: 76 h.

Goals:
Never before has so much information been available about how developers collaborate to create and maintain software systems. Data sources range from source code repositories to code reviews discussions, from comments on Twitter to educational videos on YouTube, from transcripts of developers' interviews to crash reports and execution logs. Presense of this data enabled the emergence of empirical software engineering, a subfield of software engineering, aiming at creation and validation of software engineering theories and assumptions through observation of software development practices. Theories validated in this way can be used to improve software development processes and artefacts, e.g., by providing developers with better tools or team managers with guidelines. To analyse this data empirical software engineering has built on and extended techniques from such data science fields such as artificial intelligence, statistics, social science and visualisation.

Learning outcomes:
Aims: After taking this class, the student should be able to

- Formulate and motivate research questions pertaining to software engineering, identify questions that can and that cannot be answered by means of empirical research
- Compare the suitability of different research designs and research methods in different scenarios; explain the relative strengths and weaknesses. Design empirical studies for different purposes (e.g., evaluating a tool, understanding a phenomenon); choose appropriate methods and defend the choice.
- Apply different empirical software engineering research methods, e.g., design interview protocols and user surveys, mine data from online repositories, and run appropriate statistical tests.
- Draw conclusions from empirical data and discuss issues that might have threatened their validity.
- Pesent results in writing.

Readings and text books:

-

Prerequisites:

- basic knowldge of statistics, readiness to read scientific papers, familiarity with modern software development, interest in software engineering research

Assessment breakdown:
20% Written final exam (minimum grade 5.0)
80% Homework assignments (minimum grade 5.0)

University: Eindhoven University of Technology (TU/e)
Department: Department of Mathematics and Computer Science
Course ID: 2IMI00
Course name: Seminar Process Analytics
Name and email address of the instructors: dr. Dirk Fahland (d.fahland@tue.nl)
Web page of the course:
https://tue.osiris-student.nl/#/onderwijscatalogus/extern/cursus?cursuscode=2IM100&collegejaar=2022&taal=en
Semester: 3 (Q2)
Number of ECTS: 5

Course breakdown and hours:

- Lectures: 32 h.
- Instructions: 32 h.
- Self-Study: 76 h.

Goals:
Organizations are constantly trying to improve the way processes are performed including administrative processes, logistics processes, or knowledge-intensive processes. Typically, such operational processes are supported by information systems that record events as they take place in real life in so-called event logs. Process Mining techniques allow for extracting information from event logs to discover models describing processes, organizations, and products, detect deviations, or predict outcomes and performance of processes. Unlike classical data mining techniques the objective of process mining is to describe and predict a large number of features in a single model once such as sequencing, choices, parallelism and loops of multiple activities, involvement of resources in activities, or the influence of data on the outcomes of activities. In this seminar, we will study and discuss the state-of-the-art research on Process Mining in the context of practically relevant uses cases from industry. We will discuss works along the entire process mining life-cycle: event data extraction and event log construction, various event data pre-processing methods, process model discovery, conformance checking and deviation detection, prediction based on models learned from event data, and visualization techniques. This provide students a good insight into this field of research in preparation for a Master project in the AIS group.

Students will study, present, and discuss selected research articles and book chapters, as well as Master theses. Through this material, students get an overview on the different types of problems faced in process mining from a practical perspective as well as the various research approaches to address them. Through presenting and discussing articles, students will learn to evaluate research approaches and evaluation techniques regarding the research problem and reproducibility. Through presenting and discussing previous Master theses, and by reproducing an evaluation from a scientific paper using process mining software and tools, students gain an understanding of "patterns" and "anti-patterns" in the execution of a Master project and in writing a thesis.

Students enrolling in the seminar are invited to propose their own literature and topics in preparation for their Master project, or to pick from a broad list of available topics and literature made available before the start of the seminar.

Learning outcomes:
This seminar combines teaching research methods (in preparation for a Master project) with providing students with recent and ongoing research in the area of event data analysis and process analysis. We study recent research articles, book chapters, and Master theses on topics along the entire analysis life-cycle. Through presentation and group discussions, we work out how to approach, execute, evaluate, and discuss research questions. The aim of the seminar is to prepare students for their graduation project. The specific literature and topics studied in the seminar will be made available before the first meeting; students may propose their own literature and topics in preparation for their Master project. At the end of this seminar, students:

- Have an overview on the state of the art in research (general research problems, techniques, and practically relevant use cases) in process mining and event data analysis
- Are able to independently study, present, discuss, compare, and criticize scientific texts
- Can discuss and select common research methods and techniques for a chosen problem
- Are able to propose and critically discuss research questions and experimental setups to test research questions
- Are able to conduct an experimental evaluation and discuss its validity regarding conclusions and reproducibility

Readings and text books:

- Various articles made available through the LMS/the university library

Prerequisites:

- Knowledge of process mining, data mining, data engineering, or process modeling (through relevant courses)

Assessment breakdown:
70% Written report
30% Presentation

University: CentraleSupélec, Paris-Saclay University
Department: Computer Science Department
Course ID: DeM
Course name: Decision Modeling
Name and email address of the instructors: Prof. Brice Mayag (brice.mayag@dauphine.fr)
Web page of the course: (to be created)
Semester: 3
Number of ECTS: 5

Course breakdown and hours:

- Lectures: 24h
- Laboratory: 24h
- Project: 12h

Goals:
This course aims at presenting classical decision models with a special emphasis on decision making in uncertain situations, decision with multiple attribute, and decision with multiple stakeholders. During the course, various applications will be presented, emphasizing the practical interest and applicability of the models in real-world decision situations.

Learning outcomes:
Upon successful completion of this course, the student will acquire knowledge and skills about:

- decision models, validity of the decision models
- the three levels of decision analysis: representation of observed decision behavior (descriptive analysis), decision aiding and recommendation (prescriptive analysis), and the design of artificial decision agents (normative analysis).

Readings and text books:

- Denis Bouyssou, Thierry Marchant, Marc Pirlot, Alexis Tsoukiás, Philippe Vincke, *Evaluation and decision models with multiple criteria: Stepping stones for the analyst*, Springer, International Series in Operations Research and Management Science Volume 86, 2006.
- William W. Cooper, Lawrence M. Seiford, and Kaoru Tone, *Introduction to Data Envelopment Analysis and Its Uses*, Springer, 2006.

Prerequisites:

- Operational research algorithms foundations.

Table of contents:

- Data envelopment Analysis: analysis of the efficiency of the production units.
- Decision under uncertainty and decision trees: theory, modeling and applications.
- Behavioural decision analysis: empirical analysis of decision behaviour, cognitive decision biases, prospect theory.
- Outranking methods (theory and applications): presentation of the Electre methods (Electre I, Electre 3, Electre Tri), reference based ranking.
- Applications on a generic Decision platform: decision Deck. Analysis of some use cases and use of an open source platform for decision aid.
- Group decision: group decision, elicitation of a group decision model.
- Preference learning: eliciting preference model for a decision maker, for several decision makers.
- Decision making using Multiple Objective Optimisation: epsilon constraint method, applications, approximation algorithms, evolutionary algorithms, and NSGA II.

Assessment breakdown:
Homework and class participation (10%), Written exam (50%), Project (40%)

University: CentraleSupélec, Paris-Saclay University
Department: Computer Science Department
Course ID: ML
Course name: Machine Learning
Name and email address of the instructors: Prof. Antoine Cornuéjols (Antoine.Cornuejols@Iri.fr)
Web page of the course: (to be created)
Semester: 3
Number of ECTS: 5

Course breakdown and hours:

- Lectures: 24h
- Laboratory: 24h
- Project: 20h

Goals:
The goal of this course is to provide the student with knowledge about supervised, unsupervised and reinforcement learning paradigms; the mathematical foundations and practices of different variants of machine learning methods.

Learning outcomes:
Upon successful completion of this course, the student will be able to:

- choose the best techniques to solve a given machine learning task;
- tune the parameters of the chosen method;
- interpret the results and compare different learning methods.

Readings and text books:

- David J. Hand, Heikki Mannila and Padhraic Smyth, *Principles of Data Mining*, The MIT Press, 2001.
- Trevor Hastie, Robert Tibshirani, Jerome Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition, Springer, 2009.
- Christopher Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- Koller and Friedman, *Probabilistic Graphical Models* , MIT Press, 2009.

Prerequisites:

- Familiarity with the basic probability and linear algebra theory

Table of contents:

- Supervised, unsupervised, and reinforcement learning paradigms.
- Linear and logistic regression: gradient descent, locally weighted regression, exponential families, generalized linear models.
- Generative learning algorithms, Gaussian discriminant analysis, Kernel methods, neural networks as functions, deep neural networks.
- Feature selection, curse of dimensionality, dimensionality reduction
- Probabilistic graphical models (HMM, MRF, Bayesian Networks, Inference, Kalman filters).
- Problems: on-line learning, multi-task learning.

Assessment breakdown:
Homework and class participation (10%), Written exam (50%), Project (40%)

University: CentraleSupélec, Paris-Saclay University
Department: Computer Science Department
Course ID: VA
Course name: Visual Analytics
Name and email address of the instructors: Petra Isenberg (petra.isenberg@inria.fr)
Web page of the course: (to be created)
Semester: 3
Number of ECTS: 5

Course breakdown and hours:

- Lectures: 24h
- Laboratory: 24h
- Project: 12h

Goals:
In this course students learn how to bring together automated and human-driven data analysis approaches; including innovative aspects such as: data collection, data cleaning, basic statistics, exploratory data analysis, perception and cognition, storytelling, text analysis, and multi-dimensional data representation.

Learning outcomes:
Upon completing the course, the student will be able to:

- understand basic concepts, theories, and methodologies of Visual Analytics;
- analyse data using appropriate visual analytics thinking and techniques;
- present data using appropriate visual communication and graphical methods;
- design and implement a Visual Analytics system for supporting decision making.

Readings and text books:

- Edward Tufte, *Envisioning Information*, Graphics Press, 1990.
- Robert Spence, *Information Visualisation: Design for Interaction*, Second Edition, Prentice Hall, 2007.
- Colin Ware, *Information Visualisation: Perception for Design*, Second Edition, Morgan Kaufmann, 2004.

Prerequisites:

- Database systems
- Data mining foundations

Table of contents:

- VA fundamentals: theories, methodologies and techniques;
- Designing interactive graphics;
- Appropriate methods for different data types: Graphs, Hierarchies, Spatio-temporal data, High Multi-dimensional, Text;
- Data Analysis under Uncertainty, Visualizing and exposing uncertainty;
- VA system design practices, Dashboard design;
- Exploratory Data Analysis with Tableau/Microsoft Power BI. Data Scraping in R/Python. Data Cleaning & Wrangling using OpenRefine.

Assessment breakdown:
Homework and class participation (10%), Written exam (50%), Project (40%)

University: CentraleSupélec, Paris-Saclay University
Department: Computer Science Department
Course ID: MGMA
Course name: Massive Graph Management & Analytics
Name and email address of the instructors: Prof. Binh-Minh Bui-Xuan (buixuan@lip6.fr)
Web page of the course: (to be created)
Semester: 3
Number of ECTS: 5

Course breakdown and hours:

- Lectures: 24h
- Laboratory: 24h
- Projects: 12h

Goals:
The objectives of this course is to provide the student with knowledge about designing high-performance and scalable algorithms for massive graph analytics. The course focuses on modeling and querying massive graph data in a distributed environment, designing algorithms, complexity analysis and optimization, for massive data graph problem analytics.

Learning outcomes:
Upon successful completion of this course, the student is able to:

- model and query massive graph data in a distributed environment
- design and analyse efficient graph algorithms in real-world data-intensive applications;
- develop efficient applications using the best practices in a distributed environment (Spark, MapReduce, Neo4J, GraphX, etc.).

Readings and text books:

- Richard Brath, David Jonker, *Graph Analysis and Visualization: Discovering Business Opportunity in Linked Data*, Wiley, 2015
- Ian Robinson, Jim Webber, Emil Eifrem*Graph Databases*, O'Rreilly, 2013

Prerequisites:

- Graph theory foundations
- NoSql databases in distributed environment

Table of contents:

- Modeling massive graph data in a distributed NoSql databases
- Querying massive graph data in a distributed environment
- Graph Search, Spanning Tree, Betweenness Centrality, Community Detection, Connected Components, Minimum Spanning Tree, Anomaly Detection
- Streaming Data Analysis, Data Structures for Streaming Data
- Analyzing data Streams (e.g. Twitter) and Bioinformatics data

Assessment breakdown:
Homework and class participation (10%), Written exam (50%), Project (40%)

University: CentraleSupélec, Paris-Saclay University
Department: Computer Science Department
Course ID: BDRP
Course name: Big Data Research Project
Name and email address of the instructors: Profs. Nacéra Seghouani Bennacer & Francesca Bugiotti (nacera.bennacer@centralesupelec.fr, francesca.bugiotti@centralesupelec.fr)
Web page of the course: (to be created)
Semester: 3
Number of ECTS: 5

Course breakdown and hours:

- Lectures: 18 h.
- Seminars: 12 h.
- Projects: 24 h.

Goals:
This course aims at preparing the students for the master thesis of the 4th semester. The students will learn how to manage a research project related to massive and heterogeneous data management and analytics from scratch, working in a team, and using all the steps required in a scientific methodology. During this course the students will attend seminars in order to have a better understanding of research methodologies and to be aware of some ongoing research projects presented by researchers.

Learning outcomes:
Upon successful completion of this course, the student is able to manage a scientific project from scratch in a team and provide scale-up algorithms and a program prototype for massive data management and analysis.

Readings and text books: Scientific papers will be distributed by the lecturer and the invited speakers according to the covered topics.

Prerequisites:

- Bachelor program in computer science
- Basic programming languages such as Python, Java and C++
- Basic knowledge in distributed environments (SPARK, NoSql databases)

Table of contents: The students will choose a project proposal and follow the following steps:

- Selection of relevant related works, writing citations and bibliography according to author rights (plagiarism issues);
- Summarize and classify a selection of related works;
- Formalize a research problem using the appropriate notation;
- Provide a formalized solution studying its complexity and proving its properties, to compare with existing approaches;
- Implement a prototype program (with documentation) provide experiments with detailed evaluation;
- Write a final report, and final presentation.

Assessment breakdown:
20% team project progress, 10% intermediate presentations, 10%Final defense, 30% Final report, 30% Final prototype & results

University: CentraleSupélec, Paris-Saclay University
Department: Computer Science Department
Course ID: BIM
Course name: Business Innovation Management
Name and email address of the instructors: Mr Karim Tadrist (legal affairs of Paris-Saclay) and Invited speakers from companies
(karim.tadrist@universite-paris-saclay.fr)
Web page of the course: (to be created)
Semester: 3
Number of ECTS: 5

Course breakdown and hours:

- Lectures: 30h
- Homework: 20h

Goals:
The objectives of this course are to provide the student: (i) knowledge about intellectual and industrial properties, data protection and security in European research context, (ii) an overview about current and innovative company projects and technology needs for real data analytics and machine learning.

Learning outcomes:
Upon completing the course, the student will acquire knowledge and skills about:

- intellectual and industrial properties, data protection in European research context;
- corporate and entrepreneurship culture;
- innovative projects and technologies related to massive and real data management, analytics and machine learning.

Readings and text books:

- Scientific papers will be distributed by the course lecturer and invited speakers according to the topics covered.

Prerequisites:

- Bachelor program in computer science

Table of contents:

- Lectures: Intellectual property, Industrial property, Patent law, Author rights, Data protection (European General Data Protection Regulation) and security, Sensitive data, Legal tools for startups (incubators, valorisation).
- Various seminars led by invited speakers from companies, key BI actors, and startups.

Assessment breakdown:
Oral exam (50%), Written report (50%)