ADA
ADVANCED DATA
ANALYTICS IN BUSINESS

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

REPORT

Analysis of best practice at EU partner Universities

Master courses on Data Analytics in the EU countries ADA partners

contacts:
Prof. Dr. Mirko Savić, University of Novi Sad, savicmirko@ef.uns.ac.rs
Prof. Dr Vladan Devedžić, University of Belgrade, devedzic@gmail.com
Prof. Dr Jelena Stanković, University of Nis, jelena.stankovic@eknfak.ni.ac.rs

Project acronym:	ADA

Project full title:	Advanced Data Analytics in Business
Project No:	598829-EPP-1-2018-1-RS-EPPKA2-CBHE-JP
Funding scheme:	ERASMUS+
Project start date:	November 15, 2018
Project duration:	36 months

Abstract	This document represents an overview of master study programs in data science across several EU universities, especially from EU partner universities in ADA project.

Title of document:	Analysis of best practice at EU partner universities
Work package:	WP1 Development of a new program in Advanced Data Analytics in Business
Activity:	Analysis of best practise and comparative analysis
Last version date:	01/05/2019
File name:	Analysis of best practice at EU partner universities.docx
Number of pages:	121
Dissemination level:	Project team, Department/Faculty, Institution, Regional, National

VERSIONING AND CONTRIBUTION HISTORY

Version	Date	Revision description	Partner responsible
1.0	24/04/2019	First draft	Jason Papathanasiou (CERTH)
2.0	01/05/2019	Technical corrections	Mirko Savic (UNS)

DISCLAIMER

The European Commission support for the production of this publication does not constitute an endorsement of the contents which reflects 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.

Table of Contents

Introduction

This report is about the detailed syllabi of master programs on big data in the EU countries participating in the ADA project. It was compiled thanks to the collaboration of Professor Rashid Chelouah (EISTI), Professor Gianluca Cubadda (UNITOV) and Professor Ronald Hochreiter (WU).

University of Macedonia (Greece)

Master in Business Analytics and Data Science

This is a new course inaugurated during the 2018-2019 academic year. The main aim of this Master Course is the provision of a postgraduate level specialized knowledge in business analytics and data science in University graduates, business executives and public and private sector employees. To help executives to effectively manipulate multidimensional Big Data that flow daily to Public and Private Organizations from multiple sources using the appropriate IT, Data Analysis and Operational Research - Optimization tools. The programme aims to promote knowledge and develop research in the broader field of Business Intelligence by developing knowledge and skills at the cutting edge of the three areas of the Decision Sciences cycle (Information Systems, Statistics, Operational Research). In this way it is expected that graduates of the program, will promote the upgrading of the quality of the products and services provided through their organizations, contributing to the sustainable development targets, both in the economic, social and environmental spheres.

Course structure

The total number of ECTS units is 90 and the programme of studies is as in table 1.

Table 1

First semester required – Preliminary course	ECTS

Software Tools for Business Analytics	7,5
First semester - required courses	ECTS
Introduction to Big Data and Business Intelligence Systems	7,5
Business Analytics I-Descriptive Analytics and Introduction to Predictive Analytics	7,5
Business Analytics with Management Science models and methods – Prescriptive Analytics	7,5
Second semester - Required courses	ECTS
Advanced Predictive Analytics and Data Mining	7,5
Introduction to Data Management methods and techniques	7,5
Electives (choice of 2):	2x7,5
Business Analytics II, Advanced Statistical methods and multivariate Analysis	7,5
Decision Analysis and Optimization	7,5
Marketing and Social Media Analytics	7,5
Financial Management Analytics	7,5
Operations and Supply Chain Analytics	7,5
Web and Text Analytics	7,5
Simulation Techniques in Business Analytics	7,5
Third semester	ECTS
Master thesis	30

Detailed description of the courses

a. Software Tools for Business Analytics

i. Description

Python is a modern programming language that is particularly distinguished by its easy-to-read code and ease of use. It also has a wealth of tools that make it very useful, flexible and efficient for scientific work. R is also a modern language that is mainly used for statistical processing. The course focuses on the learning of advanced techniques of the above languages for solving network problems, performing algorithmic analysis, calculating statistics and visualizing data. It is a laboratory course and all of the software used is Free Software and Open Source Software. It is also considered preparatory, in the sense that all the software and techniques taught will be used later in the work of the remaining postgraduate courses. At the beginning, a brief introduction will be made to these language functions, which are related to the definition of variables, commands, data structures, and the configuration of the interface. Then, upon completing the course, students should be able to write Python and R code related to the topics mentioned above, ie to use the specific tools to work with tables and other data structures, visualize data and enable to represent them in properly structured charts, to implement algorithms, to reproduce networks and at the same time to visualize them and finally to be able to make extensive statistical analyses.

ii. Software

Python, Anaconda, Spyder, NetworkX, Pandas, Seaborn, matplotlib, scipy, numpy, graphviz, gnuplot, gnuplot.py, R.

iii. Syllabus

Introduction to Python I (installation, editors, variables, etc)

Introduction to Python II (data structures, functions, I/O)

Arrays and scientific programming with Python (packages: numpy, scipy)

Algorithm analysis with Python

Statistics with Python I (package: pandas)

Statistics with Python II (package: StatsModels)

Networks with Python (package: NetworkX)

Network visualization with Python (packages: matplotlib, graphviz)

Data visualization with Python (packages: matplotlib, gnuplot.py, seaborn)

Introduction to R (installation, editors, variables, etc)

Statistics with R I

Statistics with R II

Final Exam

iv. Bibliography

Haslwanter, T. “An introduction to Statistics with Python. With applications in the Life Sciences”. Springer, 2016.

Johansson, R. “Numerical Python. A practical techniques approach for Industry”. Springer, 2015.

Linge, S. and **Langtangen**, H. P. “Programming for Computations - Python. A gentle introduction to Numerical Simulations with Python”. Springer, 2016.

Rahlf, T. “*Data Visualisation with R*”. Springer, 2017.

Daróczi, G. “*Mastering Data Analysis with R*”. Packt Publishing 2015.

Documentation for packages: NetworkX, Pandas, Seaborn, matplotlib, scipy, numpy, graphviz, gnuplot.py.

b. Introduction to Big Data and Business Intelligence Systems

i. Description

The aim of the course is the theoretical and practical introduction of students to the concepts of large data, business intelligence and analytical data. After completing the course students will be able to:

• Explain modern developments in the field of large data and business intelligence.

• Report successful scenarios of exploiting large data and analytical data to modern businesses worldwide.

• Describe the basic concepts and functions of data warehouses.

• Differentiate different types of data visualizations and choose the right one.

• Describe the process, methods, and tools for predictive analytics in enterprises.

• Recognize various modern large data management technologies such as Hadoop, NoSQL, graph databases, etc.

• Understand the use of software tools to exploit analytical large business data.

The lesson will include scenarios for the use of analytical data for decision-making related to modern business problems of large corporations (e.g. Instacart, Airbnb, BNP Paribas, Zillow) in various sectors such as e-commerce, banking, tourism and real estate management.

ii. Software

Python, Jupyter Notebook

iii. Syllabus

Basic concepts of Data Analytics and Business Intelligence

Data and Databases

Data Warehouses

Data visualisations

Predictive analytics

Text mining and sentiment analysis

Web mining

Social Network Analysis

Big Data: Hadoop, NoSQL, SPARK

Semantic Web and linked data

Artificial Intelligence

Privacy and Ethics

Final Exams

iv. Bibliography

Provost, F. and Fawcett T., Data Science for Business, 2013, O'Rielly, Sebastopol, CA. Sharda, R., Delen, D., Turban, E., Business Intelligence and Analytics, Systems for Decision Support, 2014, Pearson Education, Essex, England.

c. Business Analytics I-Descriptive Analytics and Introduction to Predictive Analytics

i. Description

Business decisions are often taken under conditions of uncertainty. In the modern business environment, technological developments have facilitated the collection of large data (Big Data) that may possibly improve the decision-making process. Business Analytics refers to the ways in which businesses, non-profit institutions and governments can use this data to gain knowledge and make better decisions. The ability to efficiently use data to make quick, precise and profitable decisions is a crucial strategic asset for business. Business Analytics is basically based on quantitative and statistical methods and optimization processes to identify patterns and trends in data that ultimately lead to realistic forecasts. The aim of this course is to help students learn a variety of key statistical tools useful for summarizing and presenting past events and information. Students will learn how to convert raw data into descriptive summaries that can easily be presented and understood. It will also introduce students to the fundamental concepts of Statistical Inference, such as Parameter Evaluation and Case Control, as well as statistical tools useful in

Business Analytics, such as Correlation Analysis and Time Series Analysis. Emphasis will be placed on applications, concepts and interpretation of results rather than on theory and calculations. To implement all of the above, the SPSS statistical package will be used to help students familiarize themselves with the software and be able to perform any data analysis.

ii. Software

SPSS, Excel, R

iii. Syllabus

1. Introduction to Data Analysis and Business Analytics
2. Describing and Summarizing Data
3. Visualizing and Understanding Data
4. Data preparation-Cleaning Data and data transformations
5. Descriptive Statistical Measures-Relationships between two variables
6. Probability Distributions and Data Modeling
7. Sampling and Estimation-Creating representative and unbiased samples
8. Inferential statistics-Confidence intervals
9. Inferential statistics-Designing and Performing Hypothesis Tests
10. Chi-square Tests
11. Comparative statistics-Visualizing relationships and correlation coefficient
12. Time Series Analysis and Forecasting
13. Final Exam

iv. Bibliography

James R. Evans, Business Analytics, Pearson Education, 2016.

Camm J., Cochran J., Fry M., Ohlmann J., Anderson D., Sweeney D., Williams T., Essentials of Business Analytics, Cengage Learning, 2015.

S. Christian Albright, Wayne L. Winston, Business Analytics: Data Analysis & Decision Making, Cengage Learning, 2015

Glenn J. Myatt., Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining, Willey 2007.

d. Business Analytics with Management Science models and methods – Prescriptive Analytics

i. Description

The process of making business decisions by collecting, sorting, processing and analyzing data is not new. The variation observed in recent years relates to the nature of the data sources and the form of the data themselves, having one or more of the following characteristics: large volume, greater variety of variations, velocity, variability, veracity but also value, highlighting many challenges for decision makers. In general, Business Analytics appear on three main levels of analysis: Descriptive Analytics (Data Processing and Background Information Extraction), Predictive Analytics (Past to Develop Predictive Models) and Prescriptive Analytics (using models based on previous results to propose optimal modes - prescriptions. The course focuses on the Prescriptive Analytics. That is, it first presents a general approach to some of the most important business process modeling techniques in the context of the system approach that is the basis of Operations Research or Management Science or Optimization, that is, the Science of Decisions. At the end of the course, students and students will be able to understand the role of Management Science in managing and analyzing data, developing a decision model based on a real business situation, developing solutions that provide optimal values of measures to achieve the desires of the decision-maker, to compare alternative scenarios based on these measures and to systematically explore the structure of these solutions by analyzing in depth the system and the interactions between its components. They will also have the opportunity, on the basis of decision theory, to work on a decision-making environment where subjective thinking influences significantly the "best" decision.

ii. Software

Excel, POM/QM, IBM Optimization Studio12.8 - OPL

iii. Syllabus

Introduction: The analytics era and the role of the prescriptive component, mathematical programming in Business Analytics and the linear case.

Linear Programming (LP) models: Assumptions and basic constructing principles.

Elements of optimization using Linear Programs: The graphical solution

The algorithmic perspective, Post optimality analysis and applications of LPs

Introduction to Decisions Analysis

Decision Analysis methods and applications

Introduction to Networks: Spanning trees and shortest paths.

Flows in Networks: Algorithms, LP formulation, Max Flow/Min Cut.

Integer Programming (IP) models: Assumptions and basic constructing principles.

Elements of optimization over Integer Linear Programs: The geometric and the algorithmic perspective towards optimality.

The art of modeling using boolean variables. Integer programming applications.

Final Exam.

iv. Bibliography

Asllani A., Business Analytics with Management Science Models and Methods, Pearson Education, 2015.

Camm J., Cochran J., Fry M., Ohlmann J., Anderson D., Sweeney D., Williams T., Essentials of Business Analytics, Cengage Learning, 2015.

Drake M., The Applied Business Analytics Casebook, Pearson Education, 2014.

Anderson DR, Sweeney DJ, Williams TA, Camm JD, Cochran JJ., An Introduction to Management Science 13th - 15th ed, Cengage Learning, 2010-2018.

e. Advanced Predictive Analytics and Data Mining

i. Description

Increased activity in areas such as the Internet, e-commerce, e-business, large number of online questionnaires, etc. have significantly increased the volume and complexity of data collected and stored, increasing their importance and value in making business decisions. Data Mining refers to finding a structure in large datasets using statistical techniques, artificial intelligence, and machine learning. The purpose of data mining is that the information to be extracted and standards that will become available will contribute to the decision-making process. Predictive Analytics techniques go beyond the mere description of the data and rely on the past to make forecasts for the future. These techniques are particularly important as they make it easier for business decision makers to evaluate all possible opportunities, such as revenue, profits, market share, probability of making a sale, probability of losing a customer, etc, taking into account a number of predictive factors such as marketing costs, quality assurance procedures, number of sellers, etc. This course focuses on advanced methods of data mining and predictive analytics, systematically presenting the most important predictive modeling techniques, as well as their applications to real data management, operations, marketing etc. At the end of the course it is expected that students will be able to draw and form data sets from related sources, formulate correct research questions and design plans for them, choose the appropriate techniques of data modeling and analysis leading to extract useful knowledge standards, formulating forecasting decision making. Also to evaluate and compare the effectiveness of methods and communicate the findings of the analysis to executives of organizations and businesses.

ii. Software

R, Excel, SPSS

iii. Syllabus

Introduction – Data Exploration and Data Pre-processing

Linear Regression

Cluster Analysis (Hierarchical and *k*-means)

Dimension Reduction (Principal Components Analysis)

Dimension Reduction (Correspondence Analysis)

Logistic Regression, Discriminant Analysis

Model Evaluation (Resampling Methods) – K-Nearest Neighbors, Naïve Bayes

Model Selection, Regularization and Model Tuning

Tree-based methods: Classification and regression trees

Association Rules

Support Vector Machines

Case Studies

Final Exam

iv. Bibliography

Gareth, J., Witten, D., Hastie, T. and Tibshirani, R. An Introduction to Statistical Learning: With Applications in R. New York: Springer, 2013.

Kuhn, M., and Johnson, K. Applied Predictive Modeling. New York: Springer, 2013.

Larose, D. T., and Larose, C. D. Data mining and predictive analytics. John Wiley & Sons, 2015.

f. Introduction to Data Management methods and techniques

i. Description

The aim of the course is to introduce students to the management of different forms of data found in modern enterprises. After completing the course students will be able to:

- Create data models
- Compile and execute complex SQL queries
- They distinguish the differences between traditional data management systems and large data management systems.
- Summarize the features of Hadoop and the MapReduce programming model.
- Run programs using Hadoop.
- Differentiate the different classes of NoSQL databases and describe their characteristics.

SQL teaching will be based on Oracle's educational material and learning platform as the University of Macedonia is a member of Oracle Academy. The course requires the active participation of students who will practice the subject matter in practice and will work weekly throughout the semester. The course does not require prior knowledge in programming or databases.

ii. Software

Oracle SQL, Cloudera, MongoDB, HBase

iii. Syllabus

1. Introduction to relational databases
2. Data modeling

3. The SQL language

4. Single row functions

5. Data from multiple tables

6. Aggregate reports

7. Embedded SQL Queries

8. Hadoop and MapReduce

9. Practical application of Hadoop / MapReduce

10. Key value NoSQL stores (e.g., Amazon DynamoDB, Redis)

11. NoSQL Store Document (e.g., MongoDB, Elasticsearch) - JSON

12. Extensible NoSQL stores (such as BigTable, HBase, Cassandra)

13. Final examinations

iv. Bibliography

Oracle Academy "Database Design and Programming with SQL" [Online Course]

g. Business Analytics II, Advanced Statistical methods and multivariate Analysis

i. Description

Multivariate Analysis deals with methods of collecting, describing and analyzing a set experimental units described by many variables. The use of these methods to support decision-making has been an established and widespread tactic in business for decades. However, "traditional" multivariate methods are constantly evolving towards the management of complex data sets and large volume data. The course content includes two parts. The first part presents the basic exploratory methods of multivariable data analysis. The feature of exploratory methods is that they do not distinguish variables in dependent and independent, but the purpose of the analysis is to reveal hidden relationships, tendencies, or conflicts in order to create hypotheses. This category of methods includes factorial analysis, analysis of major components, analysis of matches and cluster analysis. In the second part of the course confirmatory methods are presented. In the context of confirmatory methods, a distinction is made between independent and dependent variables and the degree of impact of the variables of the first group on the second. This includes methods such as regression analysis, multivariate variance analysis, path analysis and structural equation models. At the end of the course, students will be able to choose the appropriate methods of analysis based on research planning, the nature of the data, and the research questions related to them. They are also able to manage and represent multivariate data through statistical processing software, as well as to perform statistical analysis. Finally, they will be able to evaluate and compare the effectiveness of the methods and to report on the findings of the analysis.

ii. Software

R, Excel, SPSS

iii. Syllabus

Introduction to Multivariate Data Analysis

Exploratory Factor Analysis & Principal Component Analysis (I)

Exploratory Factor Analysis & Principal Component Analysis (II)

Correspondence Analysis

Multiple Correspondence Analysis

Cluster Analysis: Hierarchical and Partitioning Methods (I)

Cluster Analysis: Hierarchical and Partitioning Methods (II)

Multivariate Regression Analysis and Multivariate Analysis of Variance

Discriminant Analysis and Canonical Correlation

Confirmatory Factor Analysis and Path Analysis

Introduction to Structural Equation Models

Special Topics: Missing Data, Analysis of Mixed Data

Final Exam

iv. Bibliography

Bartholomew, D. J., Steele, F., Galbraith, J., & Moustaki, I. (2008). *Analysis of multivariate social science data*. Chapman and Hall/CRC.

Husson, F., Lê, S., & Pagès, J. (2017). *Exploratory multivariate analysis by example using R*. Chapman and Hall/CRC.

Schumacker, R. E. (2015). *Using R with multivariate statistics*. Sage Publications.

h. Decision Analysis and Optimization

i. Description

Optimization has been one of the cornerstones of the field of Operational Research / Management Science, an ancestor of the field we now call Business Analytics, and in particular Prescriptive Analytics. In the age of large data, optimization is the "engine" of rapid developments in the analytical sector, as the development of algorithmic methodologies capable of responding to the needs of the industry, greatly encourages its further development. This course is a continuation of the introductory course "Administrative Science in Operational Analytical", but emphasis is given to two axes with a ratio of 2 to 1: (a) the axis that highlights the algorithmic methodologies for solving problems of mathematical programming and (b) a brief introduction to decision theory, an object that demonstrates the value of subjectivity in decision-making. In the introductory lecture of the first axis there is a concise repetition of concepts and methodologies of the obligatory course OR / MS as these concepts are keys to the next modules. To solve the problems on this axis, the ILOG CPLEX 12.8 Python API will be used. The above environment is one of the dominant and most powerful convex and integer programs. The familiarity with native Python and the numpy library is taken from the required lesson "Software Tools". The linear programming sub-section will cover the form of the simplified simplex algorithm as well as its dual simplicity that is an integral part of any modern linear problem solver. In the sub-section of inline linear programming, advanced modeling techniques using binary variables as well as the two major cutting plan methods (branch and bound, Gomory's cutting plane method) will be discussed. We will then discuss some classic combinatorial optimization problems that are extremely practical and theoretically interesting. Finally, two decomposition algorithms will be presented in individual "similar" problems, a process that solves large scale optimization. It is noted that a part of the course (1/3) will be devoted to an elementary introduction to multi-attribute programming and multi-criteria decision-making methods. In this context, classical methods of multi-attribute programming and other non-parametric approaches such as the AHP method and the DEA method will be presented.

ii. Software

POM/QM, IBM ILOG CPLEX 12.8 – Python API

iii. Syllabus

From Operational Research to Business Analytics; an “optimized” evolution. Revision of key notions (OR/MS compulsory course), Linear Programing (LP).

Introduction to the Python API of IBM ILOG CPLEX. Linear programming applications.

Optimizing linear programs in practice: The revised simplex algorithm, dual linear programming and the dual simplex algorithm.

Integer Programing (IP) and the expressive modeling capacity of the integrality condition.

On the optimization of Integer Linear Programs (ILP). The branch and bound algorithm, and Gomory’s cutting plane algorithm.

Special cases of ILPs: Traveling salesman, knapsack, set covering, set packing, vehicle routing and other combinatorial optimization problems.

Large scale optimization : The Dantzig-Wolfe decomposition algorithm and

Large scale optimization II: The Bender’s decomposition algorithm

Introduction to Multiobjective optimization – basic concepts

Goal Programming

Linear models of efficiency – DEA

AHP and extensions

Final Exam.

iv. Bibliography

H.P. Williams (2013). “Model Building in Mathematical Programing - 5th edition”, John Wiley & Sons Ltd, UK.

D. Bertsimas & J.N. Tsitsiklis (1997). “Introduction to Linear Optimization”, Athena Scientific, Massachusetts, USA.

G.L. Nemhauser & L.A. Wolsey (1988). “Integer and Combinatorial Optimization”, John Wiley & Sons Ltd, USA.

Anderson D. R., D. J. Sweeney and T. A. Williams, An introduction to Management Science: Quantitative Approaches to Decision Making, ≥13th ed, Thomson.

i. Marketing and Social Media Analytics

i. Description

In the age of big data, businesses face growing challenges in terms of processing, compiling and understanding their consumer and customer data. The marketing analytics are about identifying and using specific patterns of consumer behavior through analyzing data collected internally or externally for business to resolve strategic marketing and / or decision-making problems. Students will be able to focus on marketing strategies through the use of specific analytical tools, techniques and metrics, and develop models for evaluating corporate choices. Emphasis will be placed on digital marketing tools, and in particular on the use of metrics to evaluate regular marketing on social networks and search engines. Students will be able to develop further practical skills in using the SPSS and AMOS statistical packages. In particular, students will be familiarized with analytical tools and methods used in modern marketing departments such as Cluster Analysis, Conjoint Analysis, Principal Component Analysis, Structural Equation Modeling, Regression Analysis, Decision Trees. The structure of the course reflects the modern business needs for executives with analytical skills that can support the corporate decision-making process within the marketing strategy.

ii. Software

SPSS, AMOS, Excel

iii. Syllabus

Introduction to Marketing Analytics

Who are our customers? Marketing Segmentation and Cluster Analysis

Case study: Banking Customers Segmentation

What do customers want? New Product Development and Conjoint Analysis

Understanding Customers’ Attitudes – Principal Component Analysis

Modeling Customers’ Decision Making – Structural Equation Modeling

Online Promotion Mix: Google & Social Media Analytics

Case study: Understanding Customer Value

Case study: Bank Marketing

Case study: Customer loyalty

Case study: Revenue Management to control the booking process

Presentations

Final Exam

iv. Bibliography

Hemann, C. and Burbary, K., 2013. Digital marketing analytics: Making sense of consumer data in a digital world. Pearson Education.

Mizik N., Hanssens D. M. (2018). Handbook of Marketing Analytics: Methods and Applications in Marketing, Edward Elgar Publishing: Northampton MA.

Sorger, S., 2013. Marketing Analytics: Strategic Models and Metrics. Admiral Press.

Winston, W.L., 2014. Marketing analytics: Data-driven techniques with Microsoft Excel. John Wiley & Sons.

j. Operations and Supply Chain Analytics

i. Description

Operations and Supply Chain Analytics (O/SC-Analytics) is one of the fastest growing Business Intelligence applications. An important element of the O/SC-Analytics course is timely access to trends and measurements of key performance indicators, while recent developments in information and communication technologies have contributed to a rapid increase in data-driven decision-making. The main objective of the course is to familiarize students with tactical and strategic issues around the design and operation of supply chains, develop analytical skills to solve real problems and teach students a wide range of methods and tools for the efficient management of demand and supply networks. This course studies the key areas of decision making in the design and operation of the supply chain. Students will initially learn what data they need and how they will use this data to measure supply chain performance, such as stock levels, product availability, supplier performance, warehouse efficiency and customer service levels. On this basis, they will learn how to apply different tools and methods to analyze trends, to extract knowledge and business intelligence and to make decisions. The topics covered will be divided into the planning and management of supply chain operations, including, among other things, the supplier's analysis, capacity planning, demand and supply matching, sales and function planning, position analysis and network management, inventory management, distribution, and installation locations. Finally, by analyzing and discussing case studies, they will appreciate and receive useful insights into how to optimize the value of supply chain operations and operations, rationalize objectives and design flexible supply chains.

At the end of the course the students will:

- Learn how to optimize the supply chain processes so that they can achieve a company's strategic goal of either profitability or responsiveness.

- Understand the objectives of a supply chain, explain the impact of decisions in the supply chain on the success of a company and identify key decision areas.

- Identify the main supply chain levers and determine the key performance indicators of the supply chain

- Know how to derive knowledge from dynamic information about future demand, available production capacity and sources of supply

- Develop models for network design decisions and use optimization methods for decision-making to plan the installation and analyze relevant decisions

- Use methodologies to evaluate decisions about supply chain planning and capacity allocation under uncertainty

- Apply forecasting methods to identify trends in supply and demand.

- Familiarize themselves with tools such as: EXCEL, LINGO and MCDM software

ii. Software

EXCEL, LINGO, MCDM software

iii. Syllabus

Introduction - Syllabus - Operations management analytics

Supply chain (SC) analytics-data sources - new paradigms (ie IoT, Physical Internet, Blockchain, Social media)

SC network design analytics

Predictive analytics - Collaborative Planning Forecasting and Replenishment (CPFR) analytics

Project Management Analytics

Operations research in O / SC-Analytics

Inventory management in O / SC-Analytics

Transport models analytics

Multi Criteria Decision Aid in O / SC-analytics

Markov models in O / SC-Analytics

Game theory in O / SC-Analytics

Student presentations

Final exams

iv. Bibliography

Chopra S. and Meindl P. (2012), Supply Chain Management: Strategy, Planning and Operation, 5th Edition, Pearson Education, USA.

Feigin G. (2011). Supply Chain Planning and Analytics: The right product to the right place at the right time, Business Expert Press, New York, USA.

Mathirajan, M., Sadagopan, S., Rajendran, C., Ravindran, A., Balasubramanian, P. (2016). Analytics in Operations/Supply Chain Management. I K International Publishing House.

Ramanathan, R., Mathirajan, M. and Ravindran A.R. (2017). Big Data Analytics Using Multiple Criteria Decision-Making Models. CRC Press

Singh, S. (2016). Project Management Analytics: A Data-Driven Approach to Making Rational and Effective Project Decisions. Pearson Education, Inc.

Soluade, O. (2015). Business Analytics in Production & Operations Management: A Modular Approach. LAP LAMBERT Academic Publishing

Watson, M., Lewis, S., Cacioppi, P. and Jayaraman, J. (2012). Supply Chain Network Design: Applying Optimization and Analytics to the Global Supply Chain. FT Press.

k. Web and Text Analytics

i. Description

The aim of the course is to introduce students to the analytics using data available on the World Wide Web such as open government data, linked data, data from social media, etc. After completing the course students will be able to:

• Describe data sources on the World Wide Web.

• Collect linked data through SPARQL queries

• Analyze semantically linked Web world data through visualizations and statistical analysis

• Collect, store and analyze data from social media tools

• Apply Natural Language Processing methods to social networking data.

The course requires the active participation of students who will practice the subject matter and will work weekly throughout the semester. The course does not require prior knowledge in programming or databases.

ii. Software

Virtuoso RDF store, MongoDB, Tableau, R

iii. Syllabus

1. Data sources on the World Wide Web
2. Linked data - the RDF model
3. Statistically linked data
4. The SPARQL language
5. Collection of data from the web of linked data
6. Analytically linked data
7. Data from social media
8. Data collection via Twitter API
9. Save Twitter data
10. Detailed social networking data
11. Sentiment analysis and NLP
12. Application of NLP to network society data
13. Final exams

iv. Bibliography

DuCharme B., Learning SPARQL, Second Edition, 2013, O'Reilly

T. Heath & C. Bizer, Linked Data: Evolving the Web into a Global Data Space, 2011, http://linkeddatabook.com/editions/1.0/

I. Simulation Techniques in Business Analytics

i. Description

The course focuses on simulation as one of the most popular Operational Research techniques for decision making in a non-analytical environment. Firstly, the basic theoretical aspects of the technique are presented, and then it is attempted to deepen the applications and the problems encountered with the use of software. At the end of the course, students and students will be able to develop an elementary simulation model that will describe a real problem by identifying the important elements that can influence the decision making on the basis of the objectives set and implement a systematic methodology for identifying and evaluating alternatives to the problem.

ii. Software

Excel, Extend, Simul8

iii. Syllabus

Stochastic systems and Queuing

Introduction to Simulation Modeling.

Random number generators and random variates.

Probability Distributions and input data analysis

Output Data Analysis

Discrete event simulation basics.

Simulation Techniques using Discrete event simulation environment

Applications of Discrete event simulation

Simulation with built-in Excel tools

Financial Models

Process Models

Marketing Models

Final Exam.

iv. Bibliography

Laguna M. and J. Marklund, Business Process Modeling, Simulation and Design, 2n ed 2013.

Albright S.C. and Winston W., Business Analytics: Data Analysis and Decision Making, Cengage Learning, 2013.

Extend Software, manual and reference.

Vienna University of Economics and Business (WU, Austria)

MSc in Data Science

The innovative Master's program in Data Science is all about how data can be used effectively, professionally and responsibly to gain knowledge - a complex of questions that is not only highly relevant for companies, governments and other organizations, but also for individuals. In addition, collection, modeling, analysis and interpretation of data is central to science at universities and other research institutions.

The expertise needed to address the issues involved and to develop appropriate solutions, is the main idea of the curriculum of the Master's program "Data Science", therefore goes far beyond traditional statistics and database processing. Today's Data Scientists are required holistically examine, critically scrutinize large, sometimes very heterogeneous data sources, analyze them with problem-adequate statistical methods, extract relevant information and correctly interpret results obtained.

At a technical level, not least, the enormous amount of data requires an understanding and dealing with large and often distributed systems for data storage and processing. Content development encompasses a wide range of activities, from finding and organizing the data to evaluating their quality and exploratory data analysis to modeling, analyzing and interpreting, as well as providing a clear presentation of the results. The MSc Data Science curriculum reflects the whole process chain from raw data to information, from information to knowledge and from knowledge to making informed decisions.

Structure of the MSc Program

The master program Data Science consists of 3 module groups, for which 83 ECTS credit points are planned. Furthermore, 12 ECTS credits are allocated to electives. The Master's thesis is assessed with 20 ECTS credits, the Master's examination with 2 ECTS credits and a compulsory practice with 4 ECTS credits.

	ECTS
Bridging module	12
Data Science compulsory Module	46
Data Science elective Module	24
Sum module groups	*82*
Free electives	12
Master's thesis	20
Master's examination	2
Mandatory practice (2 weeks)	4
Total	**120**

Module Content

Bridging Module

There are three different types of Bridging Modules: BM1 for Students of Mathematics, BM2 for Students of Computer Science and BM3 for Students of all other areas.

Bridging Module 1 (for Students of Mathematics)

	ECTS	I	II	III	IV
Introduction to Data Science	1	1			
Introduction to Computer Science	2	2			
Algorithms and Data Structures	4		4		
Algorithms and Data Structures - Hands-On	4		4		
Data Engineering	2		2		
Data Engineering - Hands-On	2		2		
Object-oriented Programming	2		2		
Advanced Data Engineering	2			2	
Advanced Data Engineering - Hands-On	2			2	

Bridging Module 2 (for Students of Computer Science)

	ECTS	I	II	III	IV
Introduction to Data Science	1	1			
Statistics	2	2			
Statistics - Lab	2	2			
Probability Theory	4		4		
Probability Theory - Lab	3		3		
Mathematical Statistics	3			3	
Mathematical Statistics - Lab	2			2	
Applied Statistics	3		3		

Bridging Module 3 (general)

	ECTS	I	II	III	IV
Introduction to Data Science	1	1			
Introduction to Computer Science	2	2			
Algorithms and Data Structures	4		4		
Algorithms and Data Structures - Hands-On	4		4		
Data Engineering	2		2		
Data Engineering - Hands-On	2		2		
Object-oriented Programming	2		2		
Advanced Data Engineering	2			2	
Advanced Data Engineering - Hands-On	2			2	
Statistics	2	2			
Statistics - Lab	2	2			
Probability Theory	4		4		
Probability Theory - Lab	3		3		

Mathematical Statistics	3			3	
Mathematical Statistics - Lab	2			2	
Applied Statistics	3		3		

Data Science compulsory Module

Module Statistical Methods	**ECTS**	**I**	**II**	**III**	**IV**
Statistics, Visualization and More Using R	4		4		
Computational Statistics	3			3	
Computational Statistics - Lab	3			3	

Module Databases	**ECTS**	**I**	**II**	**III**	**IV**
Database Masterclass	2			2	
Database Masterclass - Lab	3			3	
NoSQL Databases	2				2
NoSQL Databases	3				3

Module Knowledge Discovery	**ECTS**	**I**	**II**	**III**	**IV**
Machine Learning	2		2		
Machine Learning - Lab	3		3		
Pattern Recognition	2		2		
Pattern Recognition - Lab	3			3	
Data Mining	3			3	

Module Statistical Practice and Case Studies	**ECTS**	**I**	**II**	**III**	**IV**

Case Studies	4			4	
Interpreting and Presenting Statistical Analyses	4			4	

Module Law, Ethics, and Methodology of Science	**ECTS**	**I**	**II**	**III**	**IV**
Scientific Methods	4	4			
Quality of social data	4	4			
Data and Identity	2		2		

Data Science elective Module

At least two modules have to be selected - 12 ECTS are required to complete one module.

Advanced Statistical Methods and Econometrics	**ECTS**	**I**	**II**	**III**	**IV**
Econometrics	2			2	
Econometrics - Lab	4			4	
Multi-variate Statistics Masterclass	2				2
Multi-variate Statistics Masterclass - Lab	4				4

Advanced Computer Science	**ECTS**	**I**	**II**	**III**	**IV**
Advanced Algorithms	2		2		
Advanced Algorithms - Lab	4		4		
Distributed Systems	2			2	
Distributed Systems - Lab	4			4	

Parallel Programming	**ECTS**	**I**	**II**	**III**	**IV**
Parallel Algorithms	2	2			
Parallel Algorithms - Lab	4	4			
Parallel Programming	2			2	
Parallel Programming - Lab	4			4	

Image Processing and Computer Vision	**ECTS**	**I**	**II**	**III**	**IV**
Image Processing and Imaging	2	2			
Image Processing and Imaging - Lab	4	4			
Computer Vision	2			2	
Computer Vision - Lab	4			4	

Philosophy of Science	**ECTS**	**I**	**II**	**III**	**IV**
Philosophy of Science	2			2	
Philosophy of Science - Seminar	4				4
Logic I: Propositional Logic	3			3	
Logic II: Predicate Logic	3				3

Empirical Social Science & Research	**ECTS**	**I**	**II**	**III**	**IV**
Social science research methods	4	4			
Social science methodology	4		4		
Quantitative research methods	4			4	

Numerical Mathematics and Optimization	**ECTS**	**I**	**II**	**III**	**IV**
Scientific Computing	3			3	
Numerical Mathematics	3			3	
Optimization	2				2
Optimization - Lab	4				4

Module Learning Outcomes

Module - Statistical Methods - Learning Outcome

Students are able to analyze data descriptively with the aid of the statistical software R and to process results graphically (in a publishable form) as well as to draw permissible conclusions from the data. Students understand the frequentist approach to statistics, can correctly interpret errors of the first and second kind, as well as p-values and confidence intervals, and understand the possibilities and limitations of the presented methods. Students have a good overview of basic regression techniques, are able to choose parametric and non-parametric methods to suit the problem, to evaluate the quality of the fitted models, and to apply the techniques to real and simulated data using the statistical software R. On a general level, students are able to use appropriate scientific terminology of statistics.

Module - Databases - Learning Outcome

Knowledge of advanced techniques for the storage, management and retrieval of data, critical understanding of the issues of large volumes of data and complex queries, as well as an overview of the state of the art and current challenges in the field of databases. Ability to design innovative data management systems from scratch and to select and professionally use existing systems. Assessment of data management systems in terms of their capabilities and limitations for specific needs, estimating the impact of advanced data management systems on the development of computing, new services, and the organizational structure of enterprises.

Module - Knowledge Discovery - Learning Outcome

Knowledge of advanced techniques in the field of pattern recognition and machine learning, in particular their theoretical foundations, as well as the derivation of efficient algorithms in these areas. Knowledge about important libraries and software systems in these areas. Ability to apply the acquired knowledge to the analysis of existing algorithms as well as to the independent development of software for the solution of problems. Competency to select suitable libraries and / or software systems to solve practical problems with minimal self-implementation effort. Assessment of practical problems in the field of machine learning, pattern recognition as well as data mining with regard to their treatability in algorithmic and software-technical terms. Ability to assess computational complexity and select appropriate hardware.

Module - Statistical Practice and Case Studies - Learning Outcome

Students are able to analyze real data in the sense of reproducible scientific research with the help of the statistical software R, to verify clearly formulated hypotheses and to evaluate the obtained results both before and outside the subject in front of a specialized audience in English to present correctly and understandably. Based on their own analysis, students gain experience in data acquisition and preparation, descriptive and inferential analysis, model adaptation and evaluation, the use of databases and software, as well as the presentation and interpretation of results, are familiar with the respective standards and corresponding problem solving approaches. Students can reproduce simulation studies published in original articles in specialist journals and review the claimed performance. At a general level, students are able to communicate statistical findings to a specialist audience, as well as users or the general public, using appropriate scientific terminology.

Module - Law, Ethics, and Methodology of Science - Learning Outcome

Students are familiar with the relevant framework conditions for dealing with data in accordance with scientific standards with regard to legal regulations, ethical standards as well as scientific and methodological foundations. They are able to relate this knowledge to concrete questions or data. In this way, they can recognize any conflicts or difficulties in dealing with data that may arise from the aforementioned framework conditions.

Graduate school in computer science and mathematics engineering (EISTI, France)

Master Program in Advanced Data Exploration, Data Analytics and Optimization (ADEO)

Program of M1

The M1 gives the necessary bases in computer science and mathematics for the M2; we find the three pillars of which this master is characteristic. As well as the bases, students will find indispensable elements of project management. This first year will culminate in a large transversal team project.

M1 is divided into two semesters. Each semester contains 30 ECTS.

Table 4.1

Semester 1			
Skills	**Courses**	**Hours**	**ECTS**
Mathematics for Computer science	Inferential Statistics	42	12
	Partial Differential Equations and Finite Differences	30	
	Operational Research: Linear Optimization	21	
	Graph Theory and Combinatorial Optimization	21	
	Complexity and Decidability Theories	15	
Software and Architecture	Python applied to Data Science	21	10
	Object-Oriented Modeling (OOM) with UML	30	
	Object-Oriented Design and Programming with Java	30	
	Relational Database: Modeling and Design	30	

Engineering science	Signal & Information Theory	21	2
Foreign language	PPP: Personalized Professional Project	9	6
	FFL: French and Foreign languages	30	
	Total M1: Semester 1	300	30

Table 4.2

Semester 2			
Skills	**Courses**	**Hours**	**ECTS**
Data exploration	Introduction to Machine learning	24	7
	Forecasting models 1	30	
	Data analysis	21	
Mathematics for Computer Science	Deterministic and Stochastic Optimization	27	5
	Simulation and Stochastic Process	27	
Software and Architecture	Advanced database 1 (Administration, Index, Optimization)	21	9
	Architecture and Network Programming	30	
	Parallel and Distributed Programming	30	
Engineering science	Signal & Information application	30	3
Research	Research workshop	6	4
	Project	30	
Foreign	FFL: French and Foreign languages	18	2

language			
Total M1 : Semester 2		300	30

Program of M2

The M2 is, like the M1, based on the three pillars of the master, except at a higher level of expertise. To train experts in our field, we provide the students with professional skills in modelling, design and implementation of computer architecture, data mining and optimisation.

The M2 is divided into two semesters. The first one adds up to a total of 30 ECTS. The second semester is divided into two parts: The first part of the course is worth 12 ECTS and the second part consists of Master thesis with 9 ECTS and internship with 9 ECTS.

Table 4.3

Semester 1			
Skills	**Courses**	**Hours**	**ECTS**
Computer technologies	Machine learning with Scala	21	10
	Advanced data base 2 (PLSQL, Transaction, Distributed Database)	21	
	NoSQL	21	
	Dynamic web application (JEE)	21	
Data exploration	Data mining approach (Time series, logistic regression, Bagging Boosting, Random forest, Neural network)	21	7
	Semantic web and Ontology	21	
	Social Network Analysis	15	
Business Intelligent	Advanced BI & Data Visualization	24	3
Operations Research	SAS Analysis	12	7
	Forecasting models 2	33	
	Heuristics & AI	27h	

Foreign language & HR	FFL: French and Foreign Languages	26	3
	PPP: Personalized Professional Project	15	
Total M2: Semester 1		278	30

Table 4.4

Semester 2			
Skills	**Courses**	**Hours**	**ECTS**
Data exploration	Elastic search Kibana	15	3
	Text Mining and natural language	18	
	Deep learning (Convolutional Neural Network, Tensorflow, Keras,..)	12	
Operations Research	Supply Chain	18	4
	Constraint programming	18	
	Multi-objective optimization	18	
	Game theory	10	
Software and Architecture	Big data and Advanced Analytics	42	4
Foreign language	FFL: French and Foreign languages	21	1
Total courses in M2		157	12
Personal work	Master thesis		9
	Internship (22 weeks minimum)		9
Total M2 : Semester 2			30

a. Inferential Statistics

i. Objectives of the module

The objective of this course is to present the principles and the technical tools of the inferential statistics. More precisely the student will be able at the end of the course, to analyze numerical data in large quantities for the purpose to inferring proportions to a whole from those in a representative sample. We study the usual methods of estimation and tests.

The techniques introduced are illustrated during a series of tutorials by using EXCEL.

ii. Topics in details

- Reminders on Probabilities. (Random Variables Random Vectors Probability Distribution functions. Independence and Dependence of random Variables - Conditional Probabilities and Expectation values).
- Convergence. Limit theorems.

1 Estimation

- Properies of an estimator (Unbiaised Estimator-Consistent and Efficient estimator)
- Examples – Exercises)
- Usual estimators Examples – Exercises)
- Maximum Likelihood estimation Examples – Exercises).
- Estimation by interval of confidence (Examples – Exercises)

2. Hypothesis Testing.

- General principle. (Examples – Exercises)
- Test of a usual level of significance. (Examples – Exercises).
- Test of Variance. (Examples – Exercises)
- Usual tests of comparison.(One and Two samples). (Examples – Exercises)
- Chi-square tests. (Examples – Exercises)

iii. References

Arnold O. Allen « Probability Statistics and Queuing theory with Computer Science Applications” (Academic Press 1990)

Eva Cantoni, Philippe Huber, Elvezio Ronchetti : Maîtriser l’aléatoire, Springer, 2006

Kandethody M.Ramachandran, Chris P.Tsokos : Mathematical Statistics with Applications, Elsevier, 2009

George G. Roussas : A Course in Mathematical Statistics, Academic Press, 1977

G.Saporta, Probabilités Analyse des données et statistique, Editions TECHNIP

Polycopy text Tutorial By M.Manolessou

iv. Web sites

http://siba-ese.unisalento.it/index.php/ejasa/index : free access web site “Electronic Journal of Applied Statistical Analysis”

http://interstat.statjournals.net/ : free access web site “InterStat”

http://www.jds-online.com/ : free access web site “Journal of Data Science”

http://tbf.coe.wayne.edu/jmasm/ : free access web site “JOURNAL OF MODERN APPLIED STATISTICAL METHODS”

http://www.jstatsoft.org/ : free access web site “Journal of Statistical Software”

http://www.i-journals.org/ss/index.php : free access web site “Statistics Surveys”

b. Data Analysis

i. Objectives of the module

In descriptive statistics, a population is studied on one or two variables. Data analysis or multi-dimensional data analysis is an extension to several variables descriptive statistics.

This course is a first approach to the different multidimensional analysis of large masses of information methods. We discuss problems of three types: descriptive analysis, explanatory model and classification. SAS software will be used for methods on different data corpus turn.

At the end of the course, students will study for a corpus of data multidimensional:

• identify which technique to use to solve the problem;

• prepare data sets to launch the associated technical program selected;

- interpret the results provided by the software.

However, this remains an introduction course. It will take to go further :

- look more closely at the technical data preparation (very little discussed in this module);
- explore some methods discussed;
- discover new methods of analysis.

The techniques introduced are illustrated during a series of tutorials by using EXCEL

ii. Topics in details

General principles of factor analysis

Analysis of Variance (Examples exercises)

Simple Linear Regression (Examples Exercises)

Multiple Linear Regression. (Examples –Exercises)

Correlation Analysis (Montgomery and Peck Theorem) xamples –Exercises)

Non Linear regression with transformed variables (Examples and Exercises)

Principal Components Analysis

Factorial correspondence analysis

Analysis of Variance

iii. References

Arnold O. Allen « Probability Statistics and Queuing theory with Computer Science Applications” (Academic Press 1990)

Michel Volle Analyse des données Economica

G.Saporta, Probabilités Analyse des données et statistique, Editions TECHNIP

Lawley, D.N., Maxwell, A.E., Factor Analysis as a Statistical Method, Butterworths Mathematical Texts, England, 1963.

Mardia, K.V., Kent, J.T., Bibby, J.M., Multivariate Analysis, Academic Press, London 1979.

Polycopy text Tutorial by M.Manolessou

c. Introduction to Data Mining

i. Objectives of the module

This introductory course in data mining allows students to have a first approach to the problem and applications of data mining. It also allows studying several models and their applicabilities on different types of data.

iii. Topics in details

1. Data Mining fields, Data Mining Process, Data Mining Tasks, Data ant attribute natures.

2. Machine Learning: Supervised and unsupervised algorithms. Classification models, classifier validation methodology. precision and recall measures, confusion matrix, and cross validation method.

3. Comparison of supervised and unsupervised models : K-nearest neighbours and K-means algorithms

4. Supervised machine learning methods:

- Candidate elimination and version space.
- Decision Trees: ID3 and C4.5 algorithms.
- Neural Networks

5. Association Rules: Apriori and AprioriTid algorithms.

- Association rules generation.
- Properties of simple and strict redundancy

6. Comparative study and discussion.

iv. References

Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy. Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press, 1996.

Ian H. Witten; Eibe Frank Data Mining: Practical machine learning tools and techniques, 2nd Edition. Morgan Kaufmann, 2005.

Mathematics for computer science

a. Simulation and Stochastic Process

i. Objectives of the module

This course aims to study the properties of stochastic processes using simulation of random variables. It is therefore strongly practice-oriented even if the main concepts and properties are discussed

ii. Topics in details

Simulation of probability laws

Generator “random” numbers

Simulation experiments

Simulation of laws

Scheme polls to discrete distributions

Inversion method for discrete distributions

Inversion method for continuous distributions

Simulation of the normal law by the TCL

Simulation of the normal law by the Box-Muller

Stochastic Processes

Definitions and properties

Trajectories and states of a stochastic process

Properties of a process

Markov Chain

Definitions

Transient, recurrent and absorbing states

Convergence

White Noise

Definition

Simulation and validation

Brownian motion

Definition

Simulation: normalization method of random walk

Simulation: Euler method random

Validation (test of normality)

Poisson process

iii. References

- Markov models and algorithms, Bernard Ycart, Ed Springer-SMAI

b. Operational Research: Linear Optimization

i. Objectives of the module

In this course, you learn methods of linear optimization and we implement them.

ii. Topics in details

1. Linear Optimization (a) Simplexe -classical,

2. Penalties-Duality

3. Integer Numbers programming (Method of Decreasing Congruencies)

4-5. Dynamic Programming following Bellmann. Determinist cases, discrete and continuous cases. Non deterministic discrete case

6-7. Transport and Affectation problems

iii. References

G.DANTZIG``Linear programming and Extensions''Princeton, N.J.Princeton, University Press, 1963

R.FAURE ``Précis de Recherche Opérationnelle '', Dunod (Paris 1979)

S.GASS ``Linear Programming: Methods and Applications 5th edition New York : Mc Graw-Hill 1985

(Cours de l'Ecole Nationale Supérieure des Télécommunications, Paris)

C. PAPADIMITRIOU and K.STEIGLITZ ``Combinatorial Optimization: Algorithms and Complexity'' Englewood Cliffs , N.J. Prentice-Hall 1982

b. Deterministic and Stochastic Optimization

i. Objectives of the module

In this course, you learn nonlinear optimization methods and learn to implement them on a computer. Deterministic and stochastic methods and heuristic methods are addressed.

ii. Topics in details

Deterministic methods:

Gradient

Gradient with optimal step

Conjugate Gradient

Newton's method

Projection method

Method with penalty

Methods with memory/ stochastic methods :

Tabu search

Simulated annealing

Genetic algorithms

Ant colony optimization

Particles Swarm Intelligent

iii. References

- C. PAPADIMITRIOU and K.STEIGLITZ “Combinatorial Optimization: Algorithms and Complexity” (Englewood Cliffs, N.J. Prentice-Hall 1982)

- A. W. TUCKER Recent advances in Mathematical Programming (Mc GRAW-HILL, New York) W.L.WINSTON " Operations Research: Applications and Algorithms” (PWS-KENT 1991)
- I. GALEEV "Optimisation" (Science, Moscou, 2006)
- A. BJORCK "Numerical methods for least square problems" (SIAM, 1996)
- J-B. HIRIART-URRUTY, C. LEMARECHAL "Convex Analysis and Minimization Algorithms" (Springer, 1993)
- D.G.LUENBERGER " Linear and Nonlinear Programming" (Addison-Wesley, 1984)
- J. NOCEDAL, S.J. WRIGHT" Numerical Optimisation" (Springer)

c. Graph Theory and Combinatorial Optimization

i. Objectives of the module

Introduce the graph theory, and the associated algorithms.

ii. Topics in details

- Graph theory;
- Algorithms Prim, Kruskal, Dijkstra, Bellman-Ford,

iii. References

- Combinatorics and graph theory, Jean Harris, Springer Verlag, 2008.
- Modern graph theory, Bema Bollobas, Springer-Verlag New York Inc., 1st ed. 1998.
- Graph theory with applications, J.A. Bondy and U.S.R. Murty, Université of Waterloo Canada.

d. Complexity and Decidability Theories

i. Objectives of the module

Introduce the theory of decidability through its themes (a problem can be solved on a computer? Classes of problems). Give students the means to assess the difficulty of a problem and what is feasible (on computer) and what is not.

ii. Topics in details

- Turing machine;
- Formal languages;
- Decidability;
- Undecidable problems;
- Halting problem;
- Complexity classes;
- P and NP;
- NP-complete problems.

iii. References

- Computational complexity, C. H. Papadimitriou Addison-Wesley, 1994.
- Michael Sipser, Introduction to the Theory of Computation, Second Edition, Course Technology, 2005.

e. Partial Differential Equations and Finite Differences

i. Objectives of the module

In this course we will study:

- Numerical and analytical methods to solve models commonly encountered in fluid mechanics, telecommunications, biology, medicine, in industry, finance ... All these models are represented by EDP.
- Different approaches to the discretization of PDEs, stability and convergence of discrete equation. In simple cases we compare the analytical and numerical solutions.

ii. Topics in details

Course 1. Mathematical modeling and differential equations in partial derivatives

Course 2. Ordinary differential equations

Course 3. Principles of finite difference method for the PEDs

- Mesh
- Taylor formula
- Discretization of derivatives

Course 4. Basic strategy in approaches to discretization

- Explicit Euler methods
- Implicit methods Crank -Nicolson

Course 5. Boundary conditions

- Dirichlet Boundary conditions
- Neumann Boundary conditions
- Periodic Boundary conditions

Course 6. Schemes to several temporal levels

Course 7. Parabolic equations

- Thomas Algorithm
- Numerical solution of the heat equation par Crank-Nicolson. Implementation.

Course 8. Consistency, Stability. Convergence. Lax theorem

Course 9 Elliptic Equations

- Discretization of boundary conditions
- Jacobi and Gauss-Seidel iterative methods. Sparse matrix.
- Discretization and implementation in polar coordinates.

Course 10 hyperbolic equations 10.1 Advection equation

- Upwind scheme
- Lax-Friedrichs scheme

- Lax-Wendroff scheme
- Leap-Frog scheme
- Crank-Nicolson scheme

Course 11. Numerical solution of two dimensional heat equation

Course 12. Nonlinear PEDs

- Numerical solution of one dimensional Burgers equation
- Mac-Cormack method
- Crank-Nicolson method
- Numerical solution of Korteweg de Vries equation
- Numerical solution of the Sine-Gordon equation equation
- Fourier analysis of PEDs. Dispersion relation

iii. References

- H. M. Antia, Numerical Methodes for Scietists and Engineers. Birkhauser.
- M. Rappaz, M. Bellet, M. Deville, Numerical Modeling in Material Science and
- Engineering. Springer
- J.W. Thomas, Numerical Partial Differential Equations
- W.F. Ames, Numerical Methods for Partial Differential Equations, Nelson and
- Sons LTD. London, 1969
- G.D. Smith, Numerical solution of PDE : Fintite difference methods,Clarendon
- Press, Oxford, 1978
- J.H. Ferziger and M. Peric, Computational Methods for Fluid Dynamics.
- Springer, 1996.
- W. Press, S. Teokolsky, W. Vetterling, Brian P. Flannery. Numerical Recipes. The art of Scientific Computing. Cambridge University Press. 2011.
- N. Giorgano, H. Nakanishi. Computational Physics. Pearson, Pearson Hall, 2009

Engineering science

a. Signals and Systems

i. Objectives of the module

The acquisition of basic knowledge in signal processing and systems theory.

ii. Topics in details

Time Representations of Signals.

Time Representations of Systems.

Frequency Representations of Signals.

Frequency Representations of Systems

Sampling - Interpolation - Quantization.

Linear filtering. Analysis & Synthesis of digital filters. Multirate filtering.

iii. References

F. de Coulon : “ Théorie et traitement des signaux ” Dunod

P. Duvaut“ Traitement du signal ” Hermès

M. Kunt“ Traitement numérique des signaux ”Dunod

J. Max “ Méthodes et techniques de traitement du signal ”Masson

A.V. Oppenheim “ Applications of digital signal processing ” Prentice-Hall

A.V. Oppenheim / R.W. Schafer “ Digital signal processing ”Prentice-Hall

A.V. Oppenheim / A.S. Willsky / Y.T. Young “ Signals and systems” Prentice-Hall

Papoulis “ Signal analysis ”McGraw-Hill

M. Rivoire / J.L. Ferrier“ Automatique ”Eyrolles

Y. Thomas“ Signaux & systèmes linéaires ” Masson

b. Signal Processing

i. Objectives of the module

Allow the control and the design of tools for signal processing applications in the field of information processing.

ii. Topics in details

1. Mathcad simulation tool.

Tutorial.

2. Random signals. Autocovariance. Ergodicity

Transmitting a random signal in a linear system. Process for generating a random signal:

1st order formers filters. Generating process: MA, AR, ARMA

3. Signal synthesis. AR, MA, ARMA models. White noise

4. Characterization (Analysis - Frequency transforms)

- Cepstral analysis. Spectral Analysis. Wavelets. Correlation estimators.
- DSP estimator : periodogram, correlogram, from the AR model of the signal.

5. Signal conditioning. Denoising

Preaccentuation. Desaccentuation. Denoising.

6. Transmission

Coding. Equalization. Adaptive filtering.

7. Linear Prediction Coding.

Linear Prediction Coding. Lossy compression.

8. Optimal filtering.

Least squares. RLS.

9. & 10. Project

- Adaptive filtering
- Prediction (economical cycles)

- Optimal filtering
- Signal compression.
- Conditioning - Filtering. Signal detection.
- Signal characterization. Spectrogram. Wavelets.

iii. References

F. Auger“ Introduction to the signal theory ” Technip

Architecture and software

a. Object Oriented Modeling with UML

iii. Objectives of the module

This course is to learn modeling and design programs using the object approach. The language used is UML. The purpose of this course is to:

- Provide a software development methodology starting the real world until the completion of the program
- Learn how to design objects in view of reusability.

i. Topics in details

Modelling: why’s and how’s

What is in UML, what is left out ?);

’Is-kind-of’’ and ‘knows’ relations, class diagram ;

Improving models with O.C.L;

Who and which part of software is involved, when and how ? use-case diagram

Who does what in what order ? Scenarios

What means ‘state’ for objects ? state diagram

Object oriented design

Interface

Introduction to Design Patterns

iii. References

Object-Oriented Analysis and Design with Applications, 3/E ; Grady Booch, Robert A. Maksimchuk, Michael W. Engel, Bobbi J. Young, Ph.D. Jim Conallen, Kelli A. Houston ; Addison-Wesley ; 2007

Designs Patterns - Elements of Reusable Object-Oriented Software ; Erich Gamma, Richard Helm, Ralph Johnson, John Vlissides ; Addison Wesley

Object Management Group (OMG) : http://www.omg.org

b. Object Oriented Design and Programming with Java

i. Objectives of the module

This course involves learning object-oriented programming with the Java language and introduce some design patterns.

ii. Topics in details

Programmation paradigms

Classes and object members

UML Mapping of the association

Class members and inheritance

Packages

Interfaces - Eclipse

Collections

The exceptions

Input/output – Files and Flow

Generic programmation

Enumerations – Annotations

iii. References

Java in a Nutshell ; David Flanagan ; O'Reilly

http://java.sun.com

JavaDoc JRE 1.6 et 1.7

http://www.oracle.com/technetwork/java/javase/downloads/index.html#docs

c. Relational Databases Modeling and Design

i. Objectives of the module

Databases are now a central element of a vast majority of information systems, solving powerful and effective way the issue of the long-term storage of complex and extensive data.

Databases can also be regarded as an overlay file system, to provide an optimal and efficient way of storing and especially to access these data.

At first, this course introduces the concept of databases and provides the first skills in modeling, design, handling and use of data models.

Then we move on to more advanced concepts:

optimal implementation of treatments on the DBMS

design of distributed databases,

safety management through roles,

optimization of queries on large data volumes..

ii.Topics in details

The basic concepts

Entity-Relationship Model (conceptual data model, logical data model and normalization) : 2 lectures

SQL Data Definition Language : 1 lecture

SQL Data Manipulation Language : 4 lectures

Index and View : 1 lecture

Transaction : 1 lecture

c. Advanced database

i. Objectives of the module

After the first course on introduction to database, we move on to more advanced concepts:

optimal implementation of treatments on the DBMS

PLSQL

design of distributed databases,

safety management through roles,

ii. Topics in details

The advanced concepts

Processing in the database

The langague PL/SQL

the triggers

the procedures, fonctions and packages

distributed database

the concept : single MCD et multiple MLD

the different kinds of fragmentation: horizontal, vertical and mixed

the reconstituting views

the materialized views

the data bases links

Roles in a database

the application roles

the other roles

the system privileges

the object privileges

Introduction à l'administration d'une base de données

l'architecture SPARC

the tablespace

the repository

the accelerators

the indexs : b-arbres, bitmaps, inversed

the clusters

the request plan

iii. References

Oracle PL/SQL Programming de Steven Feuerstein et Bill Pribyl chez O'Reilly

d. Architecture and Network Programming

i. Objectives of the module

Discovery and familiarity with the concepts and techniques of networks.

The first part leads to developments in Java, the second part can develop in C.

We will introduce the programming R.M.I. (Java) that is widely used in the parallel computing..

ii. Topics in details

Network models

TCP Implementation in JAVA language

UDP Implementation in JAVA language

http Implementation in JAVA language

Proxy et de firewall

R.M.I. technics in java language

Network administration protocol

iii. References

TCP/IP: architecture, protocoles, applications. : Douglas Comer : InterEditions 1992

e. Parallel programming

i. Objectives of the module

- Introduce general techniques and specific algorithms of parallel and distributed computing.

- Discover new concepts related to cloud computing.

ii. Topics in details

General concepts

Multithreading programming with Java

Multiprocessing programming with java

Review the independence of loops

Limiting threads

Different modes of parallelization

Taxonomy Flynn

Complexity and Amdhal law

OpenMp

MPI

iii. References

- OpenMP official web site http://www.openmp.org

Project management

a. V Model and AGILE Methods

i. Objectives of the module

The objective of the course is to explain the two main methods of project management used today in software development projects: the V cycle and AGILE methodologies.

ii. Topics in details

From V model to "Agile" method

The manifest and the le panorama of Agile methods

SCRUM

XP

Kanban

iii. References

Extreme programming pocket guide

Agiles services and processes: Thierry Chamfrault et Claude Durand

Balancing Agility and Discipline de Guide for the PerplexedDe Barry Boehm et Richard Turner chez Addison Wesley

Scrum : le guide pratique de la méthode agile la plus populaire de Claude Aubry chez Dunod

FLE Beginners

a. FLE

i. Objectives of the module

Reach a level that allows good communication in everyday life (academic and professional life).

ii. Topics in details

French every day.

Discover the key aspects of French culture to facilitate integration.

Working from texts and audio-visual materials

Use the language from their personal experiences in the field.

Reaching the general vocabulary that related to life in school.

Systematically acquired with spots that reflect the four skills.

The year is punctuated level tests to monitor the smooth progress of the student.

Research

a. Initiation to the research

i. Objectives of the module

The objective of this module is to introduce to methodologies to analyze the scientific documents related to one of the three pillars of the master. This allows the student to prepare for the research project completed.

ii. Topics in details

Provide students scientific papers relating to courses seen during the current semester (Parallel Architecture, Data mining, Networking, ...).

The student will choose one paper and will make scientific critique as if he is a reviewer:

Present the authors problems,

How the authors are modeled theirs problems,

The methods chosen by the authors to solve their problems.

How the authors interpreted their results.

Have they presented perspectives to their work?

The bibliography is it recent? well adapted to their studies?, etc ...

b. Transverse Project and finalized research

i. Objectives of the module

The objective of this module is to confront the students a large project in working conditions in the workplace. It is a development of decision support software around the problematics of Big DATA.

At the end of this project, students will have a real experience of an IT project:

in project management in project management;

in unit testing and integration testing;

in project management

ii. Topics in details

From the required the student following these steps:

- Conduct interviews

- Detailed Functional Specifications
- Definition of the different teams and resources management project (svn, ...)
- Detailed Specifications
- Modeling and design of the database
- Design and development of unit modules
- Integration of modules in a development environment or pre-production
- Recipe and start of production

Program of M2

The M2 is, like the M1, based on the three pillars of the master, except at a higher level of expertise. To train experts in our field, we provide the students with professional skills in modelling, design and implementation of computer architecture, data mining and optimisation.

The M2 is divided into two semesters. The first one adds up to a total of 30 ECTS. The second semester is divided into two parts: The first part of the course is worth 12 ECTS and the second part consists of Master thesis with 9 ECTS and internship with 9 ECTS.

Computer technologies

a. Cloud Computing and NOSQL

i. Objective of the module

The objective of this module is to give students an understanding of the issues and challenges around NOSQL (Not Only SQL) technology and a variety of jurisdiction and implementation of certain technologies in a business context. This course is an introduction to Cloud Computing. In this course, students can learn how to make good use of Cloud Computing in Information Systems.

ii. Topic in detail

Overview of Cloud Computing, Origins and definitions, Advantages and disadvantages

Types of Cloud: SaaS, PaaS, IaaS

The known and established Cloud Operators on the market (study of tender): Google Apps, Chrome OS, Amazon Web Services, Windows Azure coupled with Visual Studio 2010, Sales Force

Storage paradigm: Oriented column, Oriented Key/Value, Oriented document, Oriented graph

Case Study, Engine and Google Big Table: The column-oriented model, the data structure dynamic, MongoDB and BSON, the contribution of the paper-oriented organization

iii. Bibliography

Cloud Computing Journal: http://cloudcomputing.sys-con.com/

Cloud Times: http://cloudtimes.org/

Computer World: http://www.computerworld.com/s/topic/158/Cloud+Computing

Cloud Computing for beginners: http://dwachira.hubpages.com/hub/What-is-cloud-computing-A-beginners-approach

b. Java EE

i. Objective of the module

This module is an introduction to the specificities of Java EE. It aims to make students familiar with Web application development based on a robust object-oriented architecture. The student can refer to a glossary for definitions of key concepts and techniques.

ii. Topic in detail

The Java EE will be composed as follows:

Getting Started with the Java EE environment

Servlet (Facade Pattern)

JSP

MVC Architecture applied to a Java EE (Pattern MVC) project

JavaBeans and Scopes

EL / JSTL (2 slots)

Cookies

iii. Bibliography

Core Servlets and Javaserver Pages: Core Technologies**,** Marty Hall and Larry Brown, Prentice Hall PTR, 2003

Head First Servlets and JSP, Bryan Basham, Kathy Sierra and Bert Bates, O'Reilly, 2004

Advanced BI and DataViz

a. BI and Data visualization

i. Objective of the module

This course helps students to understand well the BI Architecture, currents trends, BI solutions with examples, limitations of BI, importance of data discovery & self-service data visualization, choosing the right chart, and get hands on cutting edge data visualization tools with practical exercises.

ii. Topic in detail

This course introduces in the first the concept of decision-making via a chain of decisions. At the end of this course, students must understand the fundamental differences between both the operational and decision-making points of views within a functional architecture.

After the course of decision-making theory, we teach students to effectively implement a chain of decisions by introducing them to three basic steps and their tools:

Extract, transform and load (ETL)

Representation in cube (OLAP)

Reporting

Lecture 1– BI theory

Introduction to Business Intelligence, analytics and BI market trends, BI Architecture, BI Solutions, BI model design, limitations of BI tools, importance of data discovery, choosing the right chart, and self-service data visualization.

Decision-making: Who and why? -Original concept

The principles of construction

Basic modelling

Family tools

Modelling techniques

From the operational data base to decisional data base

Current Trends

Lecture 2 - Advanced BI Design

Introduction to QlikSense best practices and practical exercises, Introduction to Tableau, best practices, practical and real-life exercises, Data warehousing, advanced databases, data warehouse architecture, ETL best practices and limits, and data management solutions & limitations

Architecture BI

BI Solutions – Example of SAP BI Suite Tools

BI model design

Limits of traditionnal BI tools

Introduction to data discovery with Qlik tools

Lecture 3 – Reporting with Qlik Sense

Quick start

Load data

Create dashboard

Transform data - Model

Advanced features

Lecture 4 and 5 : Project

Advanced BI & Data visualization project with real-life analytics dataset

iii. Bibliography

The Data Warehouse Lifecycle Toolkit (2nd ed.), Kimball Ralph, Margy Ross, Warren Thornthwaite, Joy Mundy, Bob Becker, 2008, Wiley

Mastering Data Warehouse Design Relational and Dimensional Techniques, Claudia Imhoff, Jonathan G. Geiger, Nicholas Galemmo John

Students will receive the documentation of the various tools (Qlikview, essbase, Business Objects, etc.) with which they were trained.

Data Exploration

a. Machine learning and its applications

i. Objective of the module

This course presents a detailed approach of the applications and fields concerned by data mining. We will focus on several models and the way that they are put into use on different types of data. This course consists of two parts, a theoretical part and an application part. The theoretical part provides an analytical study of symbolic statistical and connectionist learning techniques. The practical work is done on Weka. An Introduction to the issue of “Big Data” and parallel data mining will then be studied. MapReduce and Mahout Framework are used.

ii. Topic in detail

Supervised learning or unsupervised . Notions of precision and recall , apparent error , confusion matrix and cross-validation.

Methods & Techniques of supervised machine learningBayesian classifier naive .

The decision trees . Algorithms ID3 , C4.5 , Cart.

The foil & reverse lookup algorithm

The association rules : apriori algorithms and aprioriTid . Generation of association rules .

Bayesian networksDiscretization methods and variable selection

forward and backward inferences .

Law of Bayesian network.

Structure Database : linear, V or hat.

Problems of prediction and diagnosis

Regression

Bootstrap and aggregation of models

Bootstrap

Aggregation by Bagging

Agregation by Boosting

Applications

Random forest

Artificial neural network.

The SVM

Neural models

Deep learning

iii. Bibliography

Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy. Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press, 1996.

Ian H. Witten; Eibe Frank Data Mining: Practical machine learning tools and techniques, 2nd Edition. Morgan Kaufmann, 2005.

Sean Owen, Robin Anil, Ted Dunning, and Ellen Friedman. Mahout in Action. Manning Publications, 1 edition, January 2011.

b. Forecasting Models

i. Topics in details

The discussed methods are:

• Introduction to short-term time series

• Single and double moving averages

- Single and double exponential smoothing
- Estimation of Trend
- Holt Model and Holt and Winter Model

ii. Objective of the module

The purpose of this course is the study of a sequence of numeric values representing the evolution of a quantity over time (temporal or time series). Such sequences of values can be expressed mathematically in order to analyze the behaviour, usually to understand the past and to predict future behaviour (short-term forecasting).

iii. Topic in detail

The discussed methods are:

Introduction to short-term time series

Single and double moving averages

Single and double exponential smoothing

Estimation of Trend

Holt Model and Holt and Winter Model

Estimation of the seasonal variations

Time series analysis with seasonality with multiple linear regressions

The detection of seasonality by autocorrelation

Tests on prediction and autocorrelation errors

The AR, MA, ARMA and ARIMA models

The software used is EXCEL and SAS

iv. Bibliography

Statistical Methods for Forecasting Bovas Abraham , Johannes Ledolter Publisher: Wiley

c. Semantic Web and Ontology

i. Objective of the module

The purpose of this course is to introduce the field of semantic web and ontologies and their uses in knowledge representation on the web as well as in the field of information retrieval. Tools and frameworks used for practical work in this course are: Protégé, Jena and Altova (XMLSPY and SemanticWorks).

ii. Topic in detail

Motivations, Definition and cake model

Ontology, theoretical notion and construction

Ontology types: domain, application and resolution ontology

Ontology representation, formalism and languages: XML, RDF, RDF(s) and OWL

Application: SPARQL and DBPEDIA

Ontology annotation, indexation and alignment

Application: Amazon recommendation system using semantic taxonomy.

iii. Bibliography

G.Antoniou and F.V. Harmelen. A semantic web primer. MIT Press, Massachusetts Institute of Technology, 2004.

W3C Tutorials: www.w3.org/

d. Social network analysis

i. Objective of the module

In many different contexts graphs are used to model complex systems interactions; we are handling now frequently in biological networks, social networks, web graphs modelling, graphs of peer-to-peer exchanges, for example. These graphs usually have nontrivial common properties that distinguish them from random graphs. The objective of this course is to introduce the issues and analysis techniques and search for this type of graphs. We rely on the Python language and NetworkX1 library.

ii. Topic in detail

Graph representation actors, relations, and links

Example: Small world, Internet communities

Social networks analysis: Degree, proximity, prestige, betweenness centrality, Clustering coefficient, Diameter

Communities' detection models and applications: Divisive algorithms (Newman), agglomerative ones (Louvain)

New approaches for communities' detections: leaders based algorithms, genetic algorithms

Multipartite graph and communities detections

Links predictions: Films recommendation in a bipartite graph, application: Movie Lens

Big graphs visualization: software Igraph

iii. Bibliography

Du simple tracement des interactions à l'évaluation des rôles et des fonctions des membres d'une communauté en réseau: une proposition dérivée de l'analyse des réseaux sociaux, Mazzoni, ISDM – Information Sciences for Decision Making, 25, 2006, pp. 477-487 E

Social network analysis. Methods and applications, S. Wasserman, K. Faust, New York, Cambridge University Press, 1994

e. SAS Analytics & SAS Miner

i. Objective of the module

Learn how to modify data for better analysis results, build and understand predictive models such as decision trees and regression models, compare and explain complex models, generate and use score code, apply association and sequence discovery to transaction data or use other modelling tools such as rule induction, gradient boosting, and support vector machines.

ii. Topic in detail

Introduction

Accessing and Assaying Prepared Data: creating a SAS Enterprise Miner project, library, and diagram, defining a data source, exploring a data source.

Introduction to Predictive Modelling with Decision Trees: cultivating decision trees, optimizing the complexity of decision trees, understanding additional diagnostic tools, autonomous tree growth options.

Introduction to Predictive Modelling with Regressions: selecting regression inputs, optimizing regression complexity, interpreting regression models, transforming input, categorical inputs, polynomial regressions.

Introduction to Predictive Modelling with Neural Networks and Other Modelling Tools: introduction to neural network models, input selection, stopped training, other modelling tools

Model Assessment: model fit statistics, statistical graphics, adjusting for separate sampling, profit matrices.

Model Implementation: internally scored data set, score code modules.

Introduction to Pattern Discovery: cluster analysis, market basket analysis (self-study).

Special Topic: ensemble models, variable selection, categorical input consolidation, surrogate models, SAS Rapid Predictive Modeler.

Case Studies: segmenting bank customer transaction histories, association analysis of Web services data, creating a simple credit risk model from consumer loan data, predicting university enrolment management.

iii. Bibliography

Herb Edelstein discusses the usefulness of data mining, A. Beck, 1997, DS Star. Vol. 1, No. 2. Available at www.tgc.com/dsstar/

SAS Institute Inc. 2002. SAS® 9 Procedures Guide. Cary, NC: SAS Institute Inc.

SAS Institute Inc. 2002. SAS/STAT® 9 User's Guide, Volumes 1, 2, and 3. Cary, NC: SAS Institute Inc.

Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems, Weiss, S. M. and C. A. Kulikowski, 1991, San Mateo, CA, Morgan Kaufmann

f. Web Mining, Web Analytics

i. Objective of the module

This course introduces the problem of web mining and its relation to the domains of personalization, user profile discovery and collaborative filtering. We will see how to adapt the methods and the techniques of data mining in order to apply them to the different types of web data. We will study three types of data: the structure of the web, behaviour of users (user log) and page content.

ii. Topic in detail

Web data representation and modelling: content, structure and user navigation

Content representation

Modelling the sessions of user navigation

Structure representation (the web graph)

Mining the different types of web data and user actions

Clustering unsupervised algorithms

K-means and association rules algorithms

Supervised algorithms: Decision tree, neural networks

Applications: User profile detection, prediction, recommendation, personalization and collaborative filtering

iii. Bibliography

Web Data Mining, Exploring Hyperlinks, Contents, and Usage Data, Bing Liu, Springer, 2007

Artwork 3D model database indexing and classification, Pattern Recognition, Philipp-Foliguet S., Jordan M., Najman L., Cousty J., vol. 44(3):588-597, 2011

g. Text Mining and natural language

i. Objective of the module

This course aims to provide students with the concepts and techniques of text analysis and classification of large masses of information. It shows the difference between natural language processing which focuses on the linguistic analysis and the text mining, which looks at statistical analysis.

We will work with the powerful SAS Text Miner tool.

ii. Topic in detail

Data Mining and Text Mining: for whom and for what?

Words and lemmatization

Linguistic Analysis

Statistical Analysis: Words and word frequency, Themes and factorial analysis of multiple correspondence, Themes and classification, and automatic extraction of keyword, Document Classification: decision tree and neural network, Open Queries: Markov chain

iii. Bibliography

Natural Language Processing with Python: Steven Bird, Ewan Klein, Edward Loper, O'Reilly Media

The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data, Authors: Ronen Feldman, James Sanger Publisher: Cambridge University Press

Text Mining and its Applications to Intelligence, Alessandro Zanasi, CRM & KM chez WIT Press

h. Elastics search & Kibana

i. Objective of the module

ii. Topic in detail

Introduction à la Suite Elastic

ElasticSearch

Présentation générale

Architecture

Cas pratiques : Partie 1

Requêtage

Cas pratiques : Partie 2

Kibana

Présentation générale

Cas pratiques : Partie 3

Logstash

Présentation générale

Cas pratiques : Partie 4

i. Advanced Big data & Data Analytics

i. Objective of the module

This course helps students to understand well the Big Data eco-system, currents trends, highlight the Big Data challenges, allow students to build on-demand Big Data applications and show them how to solve advanced analytics problems with Big Data using cutting-edge technologies.

ii. Topics in detail

1. Introduction to Big Data, market trends, tools & technologies, why we need to analyze Big Data, highlight on advanced analytics use-case with Big Data

2. Hadoop: Introduction to Hadoop, Hadoop eco-system, hive, impala, pig, flume, kafka etc with class exercises

3. Spark: Overview, spark data frames, programming in Scala & PySpark with real-life examples & class exercises. And, Spark streaming example using Twitter & Scala

4. Spark details with concrete examples and advanced analytics exercises

5. Advanced analytics use cases with SparkML (Linear regression, Decision Tree, Artificial Neural Network, Sciket -Learn) with real-life datasets

6. Big Data project: Twitter sentiment analysis & Advanced analytics project

Operations research

a. Game Theory

i. Objective of the module

Game theory provides tools to predict, understand, and optimize the result of complex decision-making processes. The purpose of this module is to introduce students to a few simple tools and examples of implementation. Game theory is applied in various fields such as economics, marketing, transport networks, energy, biology, and pursuit-evasion games.

ii. Topic in detail

Game theory for decision-making

Introduction

Concepts of game theory, rationality, solution, utility

Static games with perfect information

Normal form games

Zero sum games

Two-player games

Multiplayer games: computation of coalitions

Prudent strategies

Dominant strategies

Nash equilibrium in pure strategies

Mixed strategies

Static games with incomplete information

Bayesian equilibriums

Dynamic games

Extensive form games: decision trees

Sub Games Perfect Nash Equilibrium (SPNE)

Differential games

Repeated games

Repeated games with finite and infinite horizon

Evolutionary Game Theory (EGT)

Concept of population

Evolutionary Stable Strategies (ESS)

Evolution process

iii. Bibliography

Games and Dynamic Games, Alain Haurie, Jacek B. Krawczyk, Georges Zaccour, World Scientific – Now Publishers Series in Business vol. 1, 2012

Decision Making using Game Theory, An Introduction for Managers, Anthony Kelly, Cambridge University Press, 2003

Differential Games, A Mathematical Theory with Applications to Warfare and Pursuit, Control and Optimization, Rufus Isaacs, John Wiley & Sons Inc, New York, 1965

Dynamic Noncooperative Game Theory, 2nd edition, Tamer Basar, Geert Jan Olsder, Classics In Applied Mathematics, CL 23, SIAM, Philadelphia, 1999

b. Constraint programming, using IBM OPL Studio

i. Objective of the module

This course presents techniques and algorithms that are used for solving constraints. The interesting problematic dealing with finding efficient and optimized algorithms according to the presented problem is treated throughout this course. Used tools are: GNU Prolog and IBM CPLEX.

iii. Topic in detail

Constraint satisfaction problem, backtrack algorithms, anticipation and smaller domain choice algorithms

Consistency algorithms, AC1, AC3 & AC4 algorithms

Constraints in Gnu Prolog. N queen, Zebra problems, Sudoku, magic series, etc.

Global constraints, Hall intervals, Scheduling

Coloration and planning problem resolutions

CPLEX and OPL applications

iii. Bibliography

Essentials of Constraint Programming, Thom Frühwirth and Slim Abdennadher, Springer, 2003

Programmation par Contraintes, the Book Edition, Annick Fron, ISBN 978-918417-00-2.

c. Multi-objective and multi-criteria optimization

i. Objective of the module

The objective of this course is to provide students with methods and tools to master modelling and identify problems such as: Scheduling, Tracking, Spanning Tree, the Travelling Salesman Problem, Assignment, Vehicle routings, etc.

ii. Topic in detail

Definitions and problems

Classification of methods

Aggregation methods

Weighted average method

Goal programming

Goal attainment

The min-max

ε-constraint

Non-aggregated methods and non-Pareto

Parallel Selection (VGA)

Using genres

The lexicographic method

Methods based on Pareto

Resolution by metaheuristics

Simulated Annealing SA

Tabu Search TS

Genetic Algorithms GA

Ant Colony Optimization ACO

Particle Swarm Optimization PSO

7. Discussion

iii. Bibliography

Evolutionary Algorithms for Solving Multi-Objective Problems (2nd ed.), Coello Coello, C. A.; Lamont, G. B.; Van Veldhuizen, D. A., 2007, Springer, ISBN 978-0-387-33254-3

Evolutionary Multiobjective Optimization. Theoretical Advances and Applications, Ajith Abraham, Lakhmi Jain and Robert Goldberg, Springer, USA, 2005, ISBN 1-85233-787-7

d. Supply chain

i. Objective of the module

The first purpose of this course is to introduce the concept of supply chain management, to present the main building blocks, the main functions, the major business processes, and the performance measures.

The second one is to provide an overview of the role of Internet technologies and e-commerce in supply chain operations

The third par is to highlight the role of stochastic models (Markov chains, queuing networks); optimization models (linear programming, heuristics, constraint programming); and simulation in supply chain planning and decision-making.

ii. Topic in detail

Building Blocks, Performance Measures, Decisions

Supply Chain Inventory Management

Mathematical Foundations of Supply Chain Solutions.

Supply Chain Planning

Supply Chain Facilities Layout

Capacity Planning

Inventory Optimization

Dynamic Routing and Scheduling

Case studies

iii Bibliography

N. Viswanadham. Analysis of Manufacturing Enterprises. Kluwer Academic Publishers.

Y. Narahari and S. Biswas. Supply Chain Management: Models and Decision Making

Ram Ganeshan and Terry P. Harrison. An Introduction to Supply Chain Management

D. Connors, D. An, S. Buckley, G. Feigin, R. Jayaraman, A. Levas, N. Nayak, R. Petrakian, R. Srinivasan. Dynamic modelling for business process reengineering. IBM Research Report 19944,

W.J. Hopp and M.L. Spearman. Factory Physics: Foundations of Manufacturing Management.

Research

a. Write a scientific paper

i. Objective of the module

The objective of this module is to introduce students to methodologies write a scientific documents related to main pillars of the master as Big data, Spark, Architecture, BI, Web analysis. This allows the student to prepare themselves for the final research project. The subject of this paper is extracted from their internship or a student can choose any other subject. This paper is to complete their Master thesis

b. Master thesis

i. Objective of the module

This course helps students to understand well the Big Data eco-system, currents trends, highlight the Big Data challenges, allow students to build on-demand Big Data applications and show them how to solve advanced analytics problems with Big Data using cutting-edge technologies.

ii. Topic in detail

1. Introduction to Big Data, market trends, tools & technologies, why we need to analyze Big Data, highlight on advanced analytics use-case with Big Data

2. Hadoop: Introduction to Hadoop, Hadoop eco-system, hive, impala, pig, flume, kafka etc with class exercises

3. Spark: Overview, spark data frames, programming in Scala & PySpark with real-life examples & class exercises. And, Spark streaming example using Twitter & Scala

4. Spark details with concrete examples and advanced analytics exercises

5. Advanced analytics use cases with SparkML (Linear regression, Decision Tree, Artificial Neural Network, Sciket -Learn) with real-life datasets

6. Big Data project: Twitter sentiment analysis & Advanced analytics project

c. Internship

This internship will be done in a Laboratory or in a company

i. Objective of the module

The internship can take place in a research laboratory or in a company. The purpose of an internship in a company is twofold, to discover the world of the company, and above all, to see how a project is managed.

The trainee will be involved in a research project, and will participate in all phases of this project, from its conception to the realization of a POC (Proof Of Concept).

University of Rome Tor Vergata (UNITOV, Italy)

Master in Big Data in Business

The Master is a one-year graduate program entirely taught in English, designed to provide the participants with the necessary scientific, managerial, and technical background to work, at the highest professional level, in the area of Big Data. Graduates will receive the skills needed to manage advanced technologies in Software Engineering, Statistics, Business and Telecommunication Engineering for the design and management of Big Data.

The program is organized over 12 months in two distinct periods. For the first 7 months, the participants will attend classes where they will acquire the necessary theoretical and methodological training. The aim of the first period is the transmission of knowledge through traditional lectures, laboratory training and seminars. Frontal lectures are organized in three terms: the first two terms are the core of the program; in the third term the student can choose between elective subjects.

In the following 5 months, the participants are required to produce a written final work (Master thesis), whose topic should be agreed with the coordinator of the program. The Master thesis could be also carried out during an internship at Italian or European companies and institutions.

Course structure

CORE COURSES
I term: 18 ECTS
During the first term students will attend the following courses:

Course	Academic discipline	Lectures	Practice and seminars	ECTS
Supervised learning	Economic Statistics	36	18	6
Unsupervised learning	Statistics	36	18	6
Data management for big data analysis	Informatics	18	9	3
Security & Privacy	Telecommunications	18	9	3

II term: 18 ECTS
During the second term students will attend the following courses:

Course	Academic discipline	Lectures	Practice and seminars	ECTS
High dimensional time series	Economic Statistics	18	9	3
Topics in machine learning	Informatics	24	12	4
Architectures and systems for big data	Informatics	18	9	3
Cloud & mobile	Telecommunication s	12	6	2
Designing communication of results	Organization and Human Resource Management	12		2
Decision making processes & models	Organization and Human Resource Management	12		2
Strategic management of results	Organization and Human Resource Management	12		2

ELECTIVE COURSES
III term: 15 ECTS
In the third term, the student should obtain 15 ECTS out of the following elective courses:

Course	Academic discipline	Lectures	Practice and seminars	ECTS

Blockchain technology and applications	Telecommunications	18	9	3
Economic complexity	Theoretical Physics, Mathematical Models and Methods	18	9	3
Fundamentals of corporate finance	Organization and Human Resource Management	18		3
Monitoring and processing for the Internet of People and Machines	Telecommunications	18	9	3
Network virtualization and softwarization	Telecommunications	18	9	3
Social media analysis	Informatics	18	9	3
Text mining and document analysis	Informatics	18	9	3

MASTER THESIS

After taking the exams, students shall produce a written final work (Master thesis), which corresponds to 9 ECTS.

Activity	Academic discipline	Practice and seminars	ECTS
Master Thesis			9

Detailed description of the program

The Master in Big Data in Business is organized around four pillars: Business, Computer Science, Networking Pillar and Statistics.

Business

With the introduction of Big Data the need for more advanced data visualization capabilities has increased. Through the Big data, organizations and firms can measure and know more about their business and directly translate that knowledge into improved decision making performance; in particular, they use big data for targeting customer-centric outcomes, tap into internal data and a provide better information. Students will learn to present visual analytical results, find relevance among millions of variable, communicate concepts and hypothesis to others, and even formulate business predictions.

Computer Science

Computing is an essential part of big data analysis. Mathematical and statistical methods, in order to be applied, must be implemented as programs and software systems executed on computer platforms in such a way to provide reliable data analysis in an efficient way. Students shall be introduced to problems, methods, and tools to effectively implement and apply data analysis algorithms and methods both in small-size and in high-performance computing environments. Specific frameworks such as dealing with text data or with data from social networks are also considered.

Networking Pillar

This pillar targets the understanding and practical operation of modern, virtualized, cloud-based networks, able to instantiate secure services on-the-fly, run them anywhere in the network and shift them transparently to different locations. The pillar will make students familiar with four subjects:

- Security and Privacy: basic concepts and their applications
- Cloud and Mobile/Edge Cloud: main architecture and operation, services and platforms
- Monitoring and Processing for the Internet of People and Machines: technologies and tools for data generation and collection, e.g. machine-generated data
- Network Virtualization and Softwarization: data center architectures and current virtualization paradigms and their impact on the service design/deployment/management chain.

Statistics

This pillar focuses on the statistical analysis of High-Dimensional Data (HDD), a framework where the number of variables is larger than the number of observations. Nowadays HDD are used to optimise processes in industry and in administrations, to analyse consumer behaviour, to forecast macroeconomic and financial variables, etc. The goal of the pillar is to endow students with concepts and methods in both supervised and unsupervised statistical learning and in the analysis of large dimensional dynamic systems.

a. Supervised learning

I. Course contents

The course provides an introduction to supervised learning, focusing on both regression and classification problems. Empirical applications will be illustrated using updated software tools.

II. Methodology

Theoretical lessons and practice using R and Matlab.

III. Syllabus

Introduction to Statistical Learning

The Linear Regression Model

Resampling Methods

Model Selection and Regularization Methods

Tree-Based Methods for Regression Problems

Moving Beyond Linearity

Classification

Tree-Based Methods for Classification Problems

Support Vector Machines

IV. Bibliography

Hastie T., Tibshirani R., and J. Friedman (2011), The Elements of Statistical Learning: Data

Mining, Inference, and Prediction, 2nd ed., Springer: New York.

Web page: https://web.stanford.edu/~hastie/ElemStatLearn/

Gareth J., Witten D, Hastie T., and R. Tibshirani (2013), An Introduction to Statistical Learning:

With Applications in R, Springer: New York.

Web page: https://www-bcf.usc.edu/~gareth/ISL/

Izenman A.J, (2013) Modern Multivariate Statistical Techniques Regression, Classification, and

Manifold Learning, 2nd ed., Springer: New York.

b. Unsupervised learning

I. Course contents

The course covers the main statistical techniques used to identify latent structures (i.e. structure not directly observable) in the data.

II. Methodology

Emphasis is on principles and specific models/techniques. Each method is introduced by examples and described in mathematical formulas. Some hours of computer laboratory give to the students the possibility to practice what they learn.

III. Syllabus

- Introduction to Unsupervised Learning

- Non model based techniques

- Model based techniques

IV. Bibliography

The course material will be made available during the course: slides, readings, datasets, supplementary materials (scripts in R etc).

Bishop C.M. (2006). Pattern Recognition and Machine Learning. Springer.

Marden J.I. (2015). Multivariate Statistics. http://stat.istics.net/Multivariate/

McLachlan G.J., Peel D. (2000). Finite Mixture Models. Wiley, New York.

Duda R.O., Hart P.E., Stork D.G. (2001). Pattern Classification. Wiley, 2nd Edition.

c. Data Management for Big Data Analysis

I. Course contents

Models for big data management and analysis.

II. Methodology

The module will be laboratory-centered. Specific data management problems will be proposed to students, organized in groups and suitably tutored, asking them to organize data and to access them programmatically.

III. Syllabus

Relational modeling of data and relational databases

The NoSQL approach to data management

Introduction to Geo and Spatial Database

IV. Bibliography

Elmasri R., Navathe S., Fundamentals of Database System, 7nd ed., Pearson.

Mysql Reference manual.

Mongodb manual.

Atzeni,Ceri,Fraternali,Paraboschi,Torlone Basi di dati - ed. McGraw-Hill 4nd edition.

Elmasri R., Navathe S., Sistemi di basi di dati – Fondamenti e Complementi, 7nd ed., Pearson.

d. Security and Privacy

I. Course contents

The course aims to introduce the student to security and privacy issues and relevant protection technologies, with specific focus on data protection and applications in the context of big data. Practical labs introducing the audience to basic vulnerability assessment and penetration testing

will complement the class.

II. Methodology

The course combines both frontal lectures (especially on goals i and iii) as well as laboratory activities (especially on part ii).

III. Syllabus

The course will specifically address a subset of security and privacy issues revolving around

infrastructure security,

data protection.

The first part of the course will mainly focus on the analysis of the security best practices and protocols, the second will provide an introduction to the emerging techniques (SMC, homomorphic encryption, etc) for the secure and private computation over protected data, with specific attention to scalable approaches.

IV. Bibliography

Lecture slides will be provided during the course, along with supplementary ad-hoc material (book chapters, scientific works, standard documents, etc) complementing the slides.

Alfred J. Menezes, Paul C. van Oorschot and Scott A. Vanstone, "Handbook of applied cryptography", available at http://www.cacr.math.uwaterloo.ca/hac/

William Stallings, “Cryptography and Network Security”, McGraw Hill Stephen Thomas, “SSL and TLS Essentials”, Wiley

e. Architectures, systems and algorithms for big data computing

I. Course contents

The main goal of the course is to introduce technologies, methodologies and algorithms to manage Big Data problems.

II. Methodology

Theoretical lessons, discussion, question and answers, demonstrations, practical sessions (hands-on practice).

III. Syllabus

Introduction to Big Data, Map-Reduce and Hadoop

Hadoop2 in practice

Beyond MapReduce

Laboratory sessions focused on Hadoop, Hive and Spark.

IV. Bibliography

Slides and references at free resources on the Web. Some of these are:

MapReduce: Simplified Data Processing on Large Clusters

The Google File System

Hive – A Petabyte Scale Data Warehouse Using Hadoop

Jure Leskovec, Anand Rajaraman, Jeff Ullman - Mining of Massive Datasets – free ebook: http://www.mmds.org

f. Cloud and Mobile

I. Course contents

The course introduces the fundamental concepts of Cloud Computing, describing the different service models (PaaS, SaaS, IaaS) and the different deployment options (public, private, hybrid). Examples of public cloud infrastructures are described with hands-on practice. The open source OpenStack platform is presented (with hands-on practice) as an example of private cloud. Finally, the concepts of Mobile Edge Computing / Fog Computing are shortly introduced

II. Methodology

The course combines both frontal lectures as well as laboratory activities.

III. Syllabus

Cloud Computing: definition and fundamental concepts.

Cloud Infrasctructures.

A short introduction to Fog Computing / Mobile Edge Computing concepts

IV. Bibliography

J. Hurwitz, M. Kaufman, F. Halper, “Cloud Services for Dummies, IBM Limited Edition”, John Wiley & Sons, Inc - http://www.ibm.com/cloud-computing/files/cloud-for-dummies.pdf

g. Decision Making Processes and Models

I. Course contents

The course aims at providing students with appropriate theoretical and methodological bases for analysing the decision making processes within and between complex organizations.

II. Methodology

The course’s topics are explained through a mix of theoretical lectures, students’ simulations and teaching cases.

III. Syllabus

Decision Making Processes. Theoretical Underpinnings

Decisions, Problems, Problem Solving

Current Avenues in Decision Making Research and Practice

IV. Bibliography

Abatecola G. (2014), “Untangling Self-Reinforcing Processes in Managerial Decision Making. Co-Evolving Heuristics?”, Management Decision, 52(2), pp. 934-949.

Abatecola G., Caputo A., Cristofaro M. (2018), “Reviewing Cognitive Distortions in Managerial Decision Making. Toward an Integrative Co-Evolutionary Framework”, Journal of Management Development, 37(5), 409-424.

Abatecola G., Mandarelli G., Poggesi S. (2013), “The Personality Factor: How Top Management Teams Make Decisions. A Literature Review”, Journal of Management and Governance, 17(4), 1073-1100.

Cristofaro, M. (2017a), “Herbert Simon’s bounded rationality: its historical evolution in management and cross-fertilizing contribution”, Journal of Management History, Vol. 23 No. 2, pp. 170-190.

Hammond J.H., Keeney S.L., Raiffa H. (1998), “The Hidden Traps in Decision Making”, Harvard Business Review, 76(5), 47-58.

Kahneman D., Lovallo D., Sibony O. (2011), “The Big Idea: Before You Make that Big Decision”, Harvard Business Review, 89(6), 50-60.

Marr B. (2015), Big Data. Using Smart Big Data Analytics and Metrics to Make Better Decisions and Improve Performance, Wiley, Chichester, UK.

Simon H.A. (1988), “Problem Formulation and Alternative Generation in the Decision Making Process”, Technical Report AIP 43, Carnegie Mellon University.

h. Designing Communication of Results

I. Course contents

The evolution of the role of top managers requires refining their presentation skills. Increasingly they are called to report to shareholders, financial analysts and in general, to stakeholders and its effectiveness in communication, the company can seize opportunities.

II. Methodology

The method used is a participant-centered-learning one, i.e. the immediate involvement of the participants, who "must" try to operate the techniques proposed by the teachers.

III. Syllabus

- Managerial communication, far from the normal one
- Preparation of the communication
- Management of the meeting
- Managerial presentation
- The storyline
- The message
- The structure
- The slides
- Public speaking

IV. Bibliography

i. High-dimensional time series analysis

I. Course contents

The course covers the basic aspects of multivariate time series analysis, with the focus on modeling and forecasting a large set of variables.

II. Methodology

Theoretical lessons and classes

III. Syllabus

Multivariate Time Series

Vector autoregressive model and their shrinkage

High dimensional covariance matrices

Factor models

IV. Bibliography

Ruey S. Tsay (2014), Multivariate Time Series Analysis with R and Financial Applications,Wiley, ISBN: 978-1118617908.

Web page for the textbook:

http://faculty.chicagobooth.edu/ruey.tsay/teaching/mtsbk/

Mohsen Pourahmadi (2013), High-Dimensional Covariance Estimation, Wiley

Stock, J.H., and M.W. Watson (2011), Dynamic Factor Models, in Clements, M.P., and D.F. Hendry (eds.) Oxford Handbook of Economic Forecasting, Oxford University Press.

Fan J., Liao Y, and Liu, H. (2016) An overview of the estimation of large covariance and precision matrices, The Econometrics Journal, Vol. 19, p C1-C32.

Bauwens, L. , Laurent, S. and Rombouts, J. V. (2006), Multivariate GARCH models: a survey. J. Appl. Econ., 21: 79-109. doi:10.1002/jae.842

Koop, G., Korobilis, D. and Pettenuzzo, D. (2018). “Bayesian Compressed Vector Autoregressions” Journal of Econometrics.

l. The strategic management of big data results

I. Course contents

This course in management practice explores what managers have to do in order to increase the likelihood of the firm’s success. The course examines how organizations achieve, sustain and renew competitive advantages.

II. Methodology

The methodology used is Participant-centered Learning, which entails active class discussion of the topic of the day. Relevant contribution is expected by the students to these discussions.

III. Syllabus

- strategic management in a complex world

- competitive advantage

- mission, vision, values of an organization; leadership skills needed;

- the strategic posture; the strategic process; objective setting;

- the external environment in which the organization operates: Porter forces, Pestel analysis, Swot analysis, scenarios;

- the internal environment: value chain , resources and capabilities;

- vertical and horizontal integration: different forms;

- international management;

- innovation and blue ocean thinking.

IV. Bibliography

Slides prepared by the instructor; Exercises prepared by the instructor.

Case: Eastman Kodak Company: funtime film , Harvard Business School case no. 9-594-111, 8 may 1995

R. M. Grant "Contemporary Strategy Analysis" – 8th Edition , Blackwell Publishing, 2011

m. Topics on Machine Learning

I. Course contents

The course provides an introduction to Machine Learning methods and to their application to mining information from datasets. It is strongly related to the courses in Supervised Learning and Unsupervised Learning. An introduction to the programming and to the Python language will be provided at the beginning of the course.

II. Methodology

The course is structured in a first theoretical part, devoted to introducing some relevant and widely adopted approaches to mining information in data, and a second practical one, aimed to apply those approaches to real datasets.

III. Syllabus

- Introduction to programming in Python

- Python packages relevant for machine learning.

- Connectionist approach to machine learning: neural networks.

- Support vector machines and kernel methods.

- Ensemble methods.

- Dimensionality reduction methods.

- Applying Python in a data science task

iv. Bibliography

The main reference for the theoretical part of the course is C.M. Bishop “Pattern Recognition and Machine Learning” Springer, 2007

The main references for the programming/practical one: A.B. Downey “Think Python” O’Reilly, 2012 (freely available at http:// http://greenteapress.com/wp/think-python/) for the introduction to Python and S. Raschka “Python Machine Learning” Packt. Publishing Ltd, 2015, for its use in Machine Learning

n. Blockchain Technology and Applications

I. Course contents

The course provides an introduction to blockchain technologies, with specific focus on industrial application scenarios. The course includes hands-on labs on permissioned blockchain platforms and simple application development examples.

II. Methodology

Theoretical lessons and practice using Multichain.

III. Syllabus

Blockchains’ overview; rules to understand when blochchains are needed and when they are pointless; examples of blockchains vs examples of blockchain-like structure which are not blockchain

Background: (lightweight) review of basic crypto tools exploited in a blockchain

Distributed ledger technologies and architectures

Consensus mechanisms for permissioned blockchains; consensus in permissionless blockchains

Scripting languages and smart contract;

Application examples, and hands-on practice using a permissioned blockchain platform

IV. Bibliography

Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction Hardcover – July 19, 2016, by Arvind Narayanan, Joseph Bonneau, Edward Felten, Andrew Miller, Steven Goldfeder.

o. Economic Complexity

I. Course contents

Understanding the basic issues concerning complex systems and predictability. Network theory network statistical properties. The statistical features of world wide bipartite network.

II. Methodology

Theoretical lessons and practice using Phyton and C.

III. Syllabus

Introduction of complex systems

Network theory

Optimization

World economy as complex system

IV. Bibliography

Mark Newman: Network: an introduction. UP Oxford

Steven Strogatz: Nonlinear Dynamics and Chaos, Westview Press

p. Fundamentals of Corporate Finance

I. Course contents

This course provides students with the essential concepts in finance, as well as operational decision-making tools. Financial managers' issues will be covered and analysed.

II. Methodology

The principle of corporate finance will be illustrated and discussed through lessons, lots of exercises and case discussions.

III. Syllabus

Introduction to main concepts of finance

Constructing cash flows

Value of money in time

Investments decisions

International Capital markets

Equity capital

Bonds and debt capital

Capital Budgeting

IV. Bibliography

Brealey, Myers, Allen : “Corporate Finance” , 2006 , 8th ed. McGraw Hill

Merrill Lynch “understanding financial reports, 2003, available at student’s office

A. Damodaran “Damodaran on valuation “ Wiley finance 2ns edition

q. Monitoring and Processing for the Internet of People and Machines

I. Course contents

The course aims to introduce the student to technologies and solutions for monitoring

infrastructures and applications.

II. Methodology

The course combines both frontal lectures and laboratory activities. Based on the students’ skills and interests, the mix of theory and practice may be adapted during the course.

III. Syllabus

introduction to Monitoring, Events, Metrics and Measurement, alerts and alert management; log collection and analysis; storing and graphing metrics; data and metrics visualization.

infrastructure monitoring: network monitoring tools, cloud monitoring (VMs and containers), security monitoring and SIEMs; examples and tools

stream monitoring and analytics: sketches, Bloom-type filters and their extensions for stream analytics, one-pass filtering, behavioral monitoring via extended state machine models.

IV. Bibliography

Lecture slides will be provided during the course, along with supplementary ad-hoc material (book chapters, scientific works, standard documents, etc) complementing the slides.

James Turnbull, "The Art of Monitoring", https://artofmonitoring.com/

r. Network Virtualization and Softwarization

I. Course contents

This course describes the evolution of computing and networking platforms, with emphasis on the requirements of a service provider. Specific attention will be devoted to Data Center architectures and to the concept of virtualization.

II. Methodology

The course combines both frontal lectures as well as laboratory activities.

III. Syllabus

Introduction to Data center networking.

Virtualization.

Virtual Data Center: storage, network and applications

SDN - Software Defined Networking, proprietary and standard solutions; OpenFlow; controllers

Virtualization in a provider network; NFV - Network Function Virtualization.

ETSI NFV model

Orchestration and Orchestrators

IV. Bibliography

Course material is provided by the instructor.

s. Social Media Analysis

I. Course contents

The course provides an introduction to the adoption of Machine Learning methods in the analysis of Social Networks, both in terms of information access technologies as well as regarding the possible analytics functions over Social Networks, e.g. profiling, emergence of communities and recommending.

II. Methodology

The methodological aspects of the course will be covered by a first theoretical part, while a more practical section of the course will concentrate on the application of models and methodologies

III. Syllabus

Short Introduction to Information Management and Retrieval in the Web

Advanced Language Processing for Social Network Analysis.

Social Media Analytics

IV. Bibliography

IR - Introduction to Information Retrieval, Christopher D. Manning, Prabhakar Raghavan and Hinrich Schutze, Cambridge University Press. 2008.

Social Media Analytics - Community Detection and Mining in Social Media, Lei Tang, Huan Liu, Morgan & Claypool Publishers, 2010.

t. Text Mining and Document Analysis

I. Course contents

The course provides an introduction to natural language processing and to its applications to text mining and document analysis. Empirical applications will be illustrated using updated software tools.

II. Methodology

Theoretical lessons and practice using CoreNLP (in Java) and NLTK (in Python).

III. Syllabus

The language: linguistic models and theories

Linguistic models and systems.

Morpholgy: Finite state automaton and transducers

Syntactic analysis with context-free grammars

Parsing with context-free grammars

Semantics

Textual Entailment Recognition

IV. Bibliography

Daniel Jurafsky and James H. Martin, SPEECH and LANGUAGE PROCESSING: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (Second Edition)

London School of Economics (UK)

MSc Data Science

http://www.lse.ac.uk/Statistics/Study/MSc-programmes/MSc-Data-Science

The MSc Data Science degree provides training in data science methods, with a focus on statistical perspectives. The student will receive a thorough grounding in theory, much of it at a high mathematical level, as well as gain practical skills of applied data science, enabling them to apply advanced methods of data science and statistics to investigate real world questions.

The core courses will provide them with comprehensive coverage of some fundamental aspects of data science, computational techniques and statistical analysis. They will then choose courses from a range of optional modules ranging from distributed computing for big data and Statistical Computing, to Financial Statistics and probabilistic methods in risk management and insurance. The programme will combine traditional lectures with computer lab sessions, in which you will work with data to complete hands-on exercises using programming tools.

The capstone project or dissertation will assess their ability to take on large-scale data-based problem solving, providing a realistic example of the challenges faced in data science settings by the kinds of organisations for which the MSc programme provides natural training.

Detailed description of the courses

a. Computer Programming

i.Availability

This course is compulsory on the MSc in Applied Social Data Science. This course is available on the MSc in Applied Social Data Science, MSc in Data Science and MSc in Human Geography and Urban Studies (Research). This course is available with permission as an outside option to students on other programmes where regulations permit.

Compulsory unit for MSc in Applied Social Data Science and MSc Data Science who will be given priority. Available with permission as an outside option to students on other programmes where regulations permit and places are available.

ii.Course content

This course introduces students to the fundamentals of computer programming as students design, write, and debug computer programs using the programming language Python and R. The course will also cover the foundations of computer languages, algorithms, functions, variables, object-orientation, scoping, and assignment.

iii. Teaching

20 hours of lectures and 15 hours of classes in the MT.

Students will learn how to design algorithms to solve problems and how to translate these algorithms into working computer programs. Students acquire skills and experience as they learn Python and R, through programming assignments with an approach that integrates projectbased learning. This course is an introduction to the fundamental concepts of programming for students who lack a formal background in the field, but will include more advanced problem-solving skills in the later stages of the course. Topics include algorithm design and program development; data types; control structures; functions and parameter passing; recursion; data structures; searching and sorting; and an introduction to the principles of object- oriented programming. The primary programming languages used in the course will be Python and R.

iv. Formative coursework

Students will be expected to produce 10 problem sets in the MT.

Type: Weekly, structured problem sets with a beginning component to be started in the staff-led lab sessions, to be completed by the student outside of class. Answers should be formatted and submitted for assessment.

v. Indicative reading

Guttag, John V. Introduction to Computation and Programming Using Python: With Application to Understanding Data. MIT Press, 2016.

Gries, Paul, Jennifer Campbell, and Jason M Montojo. Practical Programming: An Introduction to Computer Science Using Python 3. The Pragmatic Bookshelf, 2013.

Miller, Bradley N. and David L. Ranum. Problem Solving with Algorithms and Data Structures Using Python. Available online at http://interactivepython.org/runestone/static/pythonds/index.html.

Python, Intermediate and advanced documentation at https://www.python.org/doc/.

Venables, William N., David M. Smith, and the R Core Team. An Introduction to R. Available online at https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf.

Zuur, Alain, Elena N. Ieno, and Erik Meesters. A Beginner's Guide to R. Springer Science & Business Media, 2009.

vi. Assessment

Take home exam (50%) and in class assessment (50%) in the MT.

Student problem sets will be marked each week, and will provide 50% of the mark.

b. Managing and Visualising Data

i.Availability

This course is compulsory on the MSc in Data Science. This course is available on the MSc in Applied Social Data Science. This course is available with permission as an outside option to students on other programmes where regulations permit.

ii. Course content

The course consists of two parts that respectively cover data manipulation and data visualisation. The focus of the course is on the fundamental principles of data manipulation and data visualisation and hands-on exercises using Python as the main programming language and various packages used by modern data scientists. The course covers workflow management for data cleaning and preparation which is typically the most time consuming part of a data science project, as well as data analysis methods, and presentation of data analysis results using various data visualisation means.

The first five weeks focus on data manipulation which covers the basic concepts such as data types and data models, including relational and non-relational database data models and query languages. Students learn how to create data model instances, load data into them, and manipulate and query data using various application programming interfaces. The course covers data structures for scientific computing and their manipulation through the Python package NumPy, which includes manipulation of multidimensional array objects, functions for performing element-wise computations using arrays, tools for reading and writing array-based datasets to files, linear algebra operations and random number generators. The course also covers the use of high-level data structures and functions designed for working with structured or tabular data through the Python package pandas. This involves using the DataFrame, a tabular column-oriented data structure and the Series, a one-dimensional labelled array object. We cover the basic concepts of relational data models and SQL query language for creating and querying database tables as well as some noSQL database models. Students will learn how to perform data analytics in Python on data imported from various data sources, including delimiter-separated file formats such as csv and tsv files, JSON and XML files, SQL databases such as MySQL and PostgresSQL as well as NoSQL databases such as various document, key-value and graph databases.

The last five weeks focus on data visualisation starting with the elements of exploratory data analysis using various statistical plots. We discuss standard plots for univariate data analysis such as histograms, smoothed histograms using kernel-density estimators, empirical cumulative distribution functions, boxplots and violin plots. We then move on to standard plots for bivariate data analysis such as scatter plots, matrix data visualisation using cluster heat maps, seriation and spectral bi-clustering methods for reordering of rows and columns of a matrix data. We discuss data visualisation techniques for common tasks such as evaluation of the predictive performance of machine learning classifiers, data dimensionality reduction, and graph data visualisation. We explain plots for evaluation of binary classifiers such as receiver operating curve plots and precision recall plots. We explain the theoretical principles of dimensionality reduction methods used for visualisation of high-dimensional data points, starting with classical methods such as multidimensional scaling to more recent methods such as stochastic neighbour embedding. We explain the basic principles of graph data visualisation methods and different graph data layouts. The data visualisations are materialised in code using various Python packages such as matplotlib, Seaborn, scikit-learn modules for clustering, manifold learning and metrics, and graphviz and networkX libraries for graph data visualisation.

iii.Teaching

20 hours of lectures and 15 hours of computer workshops in the MT.

iv. Formative coursework

Students will be expected to produce 6 problem sets in the MT.

v. Indicative reading

Mckinney, W., Python for Data Analysis, 2nd Edition, O'Reilly 2017

Muller, A. C. and Guido, S., Introduction to Machine Learning with Python, O'Reilly, 2016

Geron, A., Hands-on Machine Learning with Scikit-Learn & TensorFlow, O'Reilly, 2017

Ramakrishnan, R. and Gehrke, J., Database Management Systems, 3rd Edition, McGraw Hill, 2002

Obe, R. and Hsu, L., PostgreSQL Up & Running, 3rd Edition, O'Reilly 2017

Robinson, I., Webber, J. and Eifrem, E., Graph Databases, 2nd Edition, O'Reilly 2015

Wickham, Hadley. Ggplot2: Elegant Graphics for Data Analysis, Springer, 2009

Murray, S., Interactive Data Visualisation for the Web, O'Reilly, 2013

Matplotlib, https://matplotlib.org

Seaborn: statistical data visualization https://seaborn.pydata.org

Sci-kit learn, Machine learning in Python, http://scikit-learn.org

vi. Assessment

Project (60%) and continuous assessment (40%) in the MT.

Four of the problem sets submitted by students weekly will be assessed (40% in total). Each problem set will have an individual mark of 10% and submission will be required in MT

Weeks 3, 6, 8 and 10. In addition, there will be a take-home exam (60%) in the form of an individual project in which they will demonstrate the ability to manage data and visualise it through effective statistical graphics using principles they have learnt on the course. This may be done by publishing the visualisation and code to a GitHub repository and GitHub pages website.

c. Data Analysis and Statistical Methods

i. Availability

This course is compulsory on the MSc in Operations Research & Analytics. This course is available on the MSc in Data Science. This course is available with permission as an outside option to students on other programmes where regulations permit.

Course not available on MSc in Statistics nor on MSc in Statistics (Financial Statistics) nor on MSc in Statistics (Social Statistics).

ii. Pre-requisites

Basic knowledge in calculus and linear algebra, as well as a first course in probability and statistics.

iii. Course content

This course will provide an introduction to methods of statistics and data analysis. The statistical software R will constitute an integral part of the course, providing hands-on experience of data analysis. The syllabus will consist of:

Part I - Introduction

- Statistical Software: R

- Data exploration and visualisation

- Probability, random variables and distribution

Part II - Tools of statistical inference

- likelihood

- estimation

- hypothesis testing

Part III - Regression Methods

- linear regression

- logistic regression

Part IV - Basic time series analysis (topics as time permits)

iv. Teaching

20 hours of lectures and 10 hours of computer workshops in the MT.

v. Formative coursework

Students will be expected to produce 5 exercises in the MT.

The bi-weekly exercises will enable students to learn how to implement statistical methods in R and provide preparation for the project. They will also ensure that students learn about the different methods of statistics and data analysis, as assessed later on the examination.

vi. Indicative reading

All of Statistics, by Larry Wasserman, Springer.

Data Analysis and Graphics using R: an Example-based Appoach, by John Maindonald an John Braun, CambridgeUniversity Press.

vii. Assessment

Exam (70%, duration: 2 hours) in the summer exam period.
Project (30%) in the MT.

d. Machine Learning and Data Mining

i. Availability

This course is compulsory on the MSc in Data Science. This course is available on the MSc in Applied Social Data Science, MSc in Marketing, MSc in Quantitative Methods for Risk Management, MSc in Statistics, MSc in Statistics (Financial Statistics), MSc in Statistics (Financial Statistics) (Research), MSc in Statistics (Research), MSc in Statistics (Social Statistics) and MSc in Statistics (Social Statistics) (Research). This course is available with permission as an outside option to students on other programmes where regulations permit.

ii. Pre-requisites

The course will be taught from a statistical perspective and students must have a solid understanding of linear regression models

Students are not permitted to take this course alongside Algorithmic Techniques for Data Mining (MA429)

iii. Course content

Machine learning and data mining are emerging fields between statistics and computer science which focus on the statistical objectives of prediction, classification and clustering and are particularly orientated to contexts where datasets are large, the so-called world of 'big data'. This course will start from the classical statistical methodology of linear regression and then build on this framework to provide an introduction to machine learning and data mining methods from a statistical perspective. Thus, machine learning will be conceived of as 'statistical learning', following the titles of the books in the essential reading list. The course will aim to cover modern non-linear methods such as spline methods, generalised additive models, decision trees, random forests, bagging, boosting and support vector machines, as well as more advanced linear approaches, such as ridge regression, the lasso, linear discriminant analysis, k-means clustering, nearest neighbours.

iv. Teaching

20 hours of lectures and 10 hours of computer workshops in the MT.

The first part of the course reviews regression methods and covers linear and quadratic discriminant analysis, cross-validation, variable selection, nearest neighbours, shrinkage, dimension reduction methods. The second part of the course introduces non-linear models and covers, splines, generalized additive models, tree methods, bagging, random forest, support vector machines, principal components analysis, k-means, hierarchical clustering.

Week 6 will be used as a reading week.

v. Formative coursework

Students will be expected to produce 5 problem sets and 1 project in the MT.

The problem sets will consist of some theory questions and data problems that require the implementation of different methods in class using a computer package.

vi. Indicative reading

James, G., Witten, D., Hastie, T. and Tibshirani, R. An Introduction to Statistical Learning. Springer, 2014. Available online at http://www-bcf.usc.edu/~gareth/ISL/

Hastie, T., Tibshirani, R. and Friedman, J. The Elements of Statistical Learning: Data Mining, Inference and Prediction. 2nd Edition, Springer, 2009. Available online at http://statweb.stanford.edu/~tibs/ElemStatLearn/index.html

Bishop, G. Pattern Recognition and Machine Learning. Springer-Verlag, 2006.

vii. Assessment

Exam (70%, duration: 2 hours) in the summer exam period.
Project (30%) in the MT Week 11.

e. Capstone Project

i. Availability

This course is compulsory on the MSc in Data Science. This course is not available as an outside option.

ii. Course content

The capstone project will provide students with the opportunity to study in depth a topic of specific interest. The topic will normally relate to a specific data source or sources and will require the use of data science skills learnt on the programme. The topic for a capstone project will be similar to that for the kinds of data-based issues faced in practice by private or public sector organisations. The capstone project is typically conducted in partnership with a company partner and is jointly supervised by the LSE faculty and company partner collaborators. The capstone project partner proposes a data science research project, potentially provides access to data, and engages through participation in joint meetings that are either online or onsite. The capstone project may require students to spend some time on company partner's premises, for example, to have access to data. The capstone project requires creative work in formulating research questions and hypotheses, identifying most suited methodology, referring to research literature, and analysing data sources using data science computing technologies.

iii. Teaching

A topic and project supervisor will be identified during MT. Supervisors will provide advice from the end of MT until two weeks after the end of ST. The student will prepare and submit project report by a date in August.

iv. Formative coursework

Formative assessment is via informal feedback from supervisors on the project report and contributions to the project as an individual contributor and team member.

Other courses on the MSc programme will also provide a range of formative assessments of relvance to the outcomes of this project.

v. Assessment

Project (100%) in August.

Maximum page limit of 50 single-sided sheets of A4 (minimum font size of 11pt and line spacing 1.5).

f. Distributed Computing for Big Data

i. Availability

This course is available on the MSc in Applied Social Data Science, MSc in Data Science and MSc in Operations Research & Analytics. This course is available with permission as an outside option to students on other programmes where regulations permit.

ii. Pre-requisites

Basic knowledge of Python or some other programming knowledge is desirable.

iii. Course content

The course covers basic principles of systems for distributed processing of big data including distributed file systems; distributed computation models such as Mapreduce, resilient distributed datasets, and distributed dataflow graph computations; structured querying over large datasets; graph data processing systems; stream data processing systems; scalable machine learning algorithms for classification, regression, collaborative filtering, topic modelling and other tasks. The course enables students to learn about the principles and gain hands-on experience in working with the state of the art big data computing technologies such as Apache Spark, a general engine for large-scale data processing, and Apache TensorFlow, a popular software library for (distributed) learning of deep neural networks. Through weekly exercises and course project work, student can gain experience in performing data analytics tasks on their laptops and cloud computing platforms. For more information, please see the course handout: http://lse-st446.github.io

iv. Teaching

20 hours of lectures and 15 hours of computer workshops in the LT.

v. Formative coursework

Students will be expected to produce 10 problem sets in the LT.

Eight of the weekly problem sets will represent formative coursework. The other two will represent summative assessment.

vi. Indicative reading

Karau, H., Konwinski, A., Wendell, P. and Zaharia, M., Learning Spark: Lightining-fast Data Analysis, O'Reilly, 2015

Karau, H. and Warren, R., High Performance Spark: Best Practices for Scaling & Optimizing Apache Spark, O'Reilly, 2017

Drabas, T. and Lee D., Learning PySpark, Packt, 2016

White, T., Hadoop: The Definitive Guide, O'Reilly, 4th Edition, 2015

Apache Spark Documentation https://spark.apache.org/docs/latest

Apache TensorFlow Documentation https://www.tensorflow.org/get_started

vii. Assessment

Project (80%) in the LT.

Continuous assessment (10%) in the MT Week 4.

Continuous assessment (10%) in the MT Week 7.

The main assessment will consist of an individual project to develop a package for fitting statistical models of the student's own choice to big data sets.

In addition, among the 10 weekly problem sets, there will be two (in weeks 4 and 7) which will contribute to summative assessment (10% each).

g. Statistical Computing

i. Availability

This course is available on the MSc in Data Science, MSc in Operations Research & Analytics, MSc in Statistics, MSc in Statistics (Financial Statistics), MSc in Statistics

(Financial Statistics) (Research), MSc in Statistics (Research), MSc in Statistics (Social Statistics) and MSc in Statistics (Social Statistics) (Research). This course is available with permission as an outside option to students on other programmes where regulations permit.

ii. Course content

An introduction to the use of numerical linear algebra, optimisation, numerical integration and simulation in statistical computation, with their applications in statistical methods, including least squares, maximum likelihood, principle component analysis, LASSO, etc. If time permits, more advanced topics such as kernel methods and graphical LASSO will also be covered. Throughout the course, students will gain practical experience of implementing these computational methods in a programming language. Learning support will be provided for at least one programming language, such as R, Python or C++, but the choice of language supported may vary between years, depending on judged benefits to students, whether in terms of pedagogy or resulting skills. This year, the default choice is Python.

iii. Teaching

20 hours of lectures and 10 hours of computer workshops in the LT.

Lectures will cover:

(1) **Introduction to Tools in Numerical Analysis**: linear algebra (Gaussian elimination, Cholesky decomposition, matrix inversion and condition); numerical optimization (bi-section, steepest descent, Newton’s method, Quasi-Newton methods, stochastic search); convex optimization (coordinate descent, ADMM); numerical integration.

(2) **Introduction to Tools in Numerical Simulation**: random number generation (inverse CDF, rejection, Box-Muller, etc); Introduction to Monte-Carlo methods.

(3) **Applications in Statistics**: linear regression and least squares; generalised linear models; principle component analysis (PCA); Page rank; LASSO.

(4) **Other advanced topics** if time allows: bootstrapping; kernel density estimation; Graphical models and Graphical LASSO.

Week 6 will be used as a reading week.

iv. Formative coursework

Students will be expected to produce 5 problem sets in the LT.

Bi-weekly exercises, involving computer programming and some theory.

v. Indicative reading

Computational Statistics by Givens and Hoeting

Statistical computing in C++ and R by Eubank and Kupresanin

The Art of R Programming: A Tour of Statistical Software Design by Matloff

Think Python: How to Think Like a Computer Scientist by Downey

vi. Assessment

Exam (70%, duration: 2 hours) in the summer exam period.
Project (30%) in the LT.

h. Multivariate Methods

i. Availability

This course is available on the MPhil/PhD in Statistics, MSc in Data Science, MSc in Marketing, MSc in Statistics, MSc in Statistics (Financial Statistics), MSc in Statistics (Financial Statistics) (Research), MSc in Statistics (Research), MSc in Statistics (Social Statistics) and MSc in Statistics (Social Statistics) (Research). This course is available with permission as an outside option to students on other programmes where regulations permit.

ii. Pre-requisites

Students must have completed Further Mathematical Methods (MA212) and Probability, Distribution Theory and Inference (ST202).

iii. Course content

An introduction to the theory and application of modern multivariate methods used in the Social Sciences: Multivariate normal distribution, principal components analysis, factor analysis, latent variable models, latent class analysis and structural equations models.

iv. Teaching

20 hours of lectures and 8 hours of computer workshops in the LT.

Week 6 will be used as a reading week.

v. Formative coursework

Coursework assigned fortnightly and returned to students via Moodle with comments/feedback before the computer workshops.

vi. Indicative reading

D J Bartholomew, F Steele, I Moustaki & J Galbraith, Analysis of Multivariate Social Science Data (2nd edition);

D J Bartholomew, M Knott & I Moustaki, Latent Variable Models and Factor Analysis: a unified approach;

C Chatfield & A J Collins, Introduction to Multivariate Analysis;

B S Everitt & G Dunn, Applied Multivariate Data Analysis;

K.V. Mardia, J.T. Kent and J.M. Bibby, Multivariate Analysis.

vii. Assessment

Exam (100%, duration: 2 hours) in the summer exam period.

i. Generalised Linear Modelling and Survival Analysis

i. Availability

This course is compulsory on the MSc in Statistics (Social Statistics) and MSc in Statistics (Social Statistics) (Research). This course is available on the MPhil/PhD in Statistics, MSc in Data Science, MSc in Marketing, MSc in Statistics, MSc in Statistics (Financial Statistics), MSc in Statistics (Financial Statistics) (Research) and MSc in Statistics (Research). This course is available with permission as an outside option to students on other programmes where regulations permit.

ii. Pre-requisites

Mathematics to the level of Mathematical Methods (MA100) and probability to the level of Probability, Distribution Theory and Inference (ST202). Some knowledge of linear regression.

iii. Course content

An introduction to the theory and application of generalised linear models for the analysis of continuous, categorical, count and survival data. Topics include: linear regression, analysis of variance (ANOVA), logistic regression for binary data, models for ordered and unordered (nominal) responses, log-linear models for count data and contingency tables, and models for survival (duration) data. The Stata software package will be used in computer workshops.

iv. Teaching

20 hours of lectures and 15 hours of computer workshops in the LT.

Week 6 will be used as a reading week.

v. Formative coursework

Coursework assigned weekly based on the computer sessions and returned to students with comments/feedback.

vi. Indicative reading

Dobson, A.J. & Barnett, A.G. (2002) An Introduction to Generalised Linear Modelling. 2nd edition. Chapman & Hall.

McCullagh, P. & Nelder, J.A. (1989) Generalized Linear Models. 2nd edition. Chapman & Hall.

Agresti, A. (2015) Foundations of Linear and Generalized Linear Models. Wiley [Available as electronic resource from LSE library].

Hosmer, D.W. & Lemeshow, S. (1999) Applied Survival Analysis, Regression Modeling of Time-to-Event Data. Wiley.

Long, J.S. and Freese, J. (2006) Regression Models for Categorical Dependent Variables Using Stata. 2nd edition. Stata Press.

vii. Assessment

Exam (100%, duration: 2 hours) in the summer exam period.

j. Time Series

i. Availability

This course is compulsory on the MSc in Statistics (Financial Statistics) and MSc in Statistics (Financial Statistics) (Research). This course is available on the MSc in Applicable Mathematics, MSc in Data Science, MSc in Econometrics and Mathematical Economics, MSc in Financial Mathematics, MSc in Marketing, MSc in Operations Research & Analytics, MSc in Quantitative Methods for Risk Management, MSc in Statistics and MSc in Statistics (Research). This course is available with permission as an outside option to students on other programmes where regulations permit.

ii. Pre-requisites

Good undergraduate knowledge of statistics and probability.

iii. Course content

A broad introduction to statistical time series analysis for postgraduates: what time series analysis can be useful for; autocorrelation; stationarity; causality; basic time series models: AR, MA, ARMA; ARCH and GARCH models for financial time series; trend removal and seasonal adjustment; invertibility; spectral analysis; estimation; forecasting. We will also discuss nonstationarity and multivariate time series if time permits.

iv. Teaching

20 hours of lectures and 10 hours of seminars in the MT.

Exercises will be given out to do at home during Week 6.

v. Formative coursework

Weekly exercises will be given.

vi. Indicative reading

Brockwell & Davis, Time Series: Theory and Methods;

Brockwell & Davis, Introduction to Time Series and Forecasting;

Box & Jenkins, Time Series Analysis, Forecasting and Control;

Shumway & Stoffer, Time Series Analysis and Its Applications.

vii. Assessment

Exam (100%, duration: 2 hours) in the summer exam period.

k. Financial Statistics

i. Availability

This course is compulsory on the MSc in Statistics (Financial Statistics) and MSc in Statistics (Financial Statistics) (Research). This course is available on the MSc in Data Science and MSc in Quantitative Methods for Risk Management. This course is available with permission as an outside option to students on other programmes where regulations permit.

ii. Pre-requisites

Students must have completed Statistical Inference: Principles, Methods and Computation (ST425) and Time Series (ST422).

iii. Course content

The course covers key statistical methods and data analytic techniques most relevant to finance. Hands-on experience in analysing financial data in the “R” environment is an essential part of the course. The course includes a selection of the following topics: obtaining financial data, low- and high-frequency financial time series, ARCH-type models for low-frequency volatilities and their simple alternatives, predicting equity indices (case study), Markowitz portfolio theory and the Capital Asset Pricing Model, machine learning in financial forecasting, Value at Risk, simple trading strategies. The course ends with an extended case study involving making predictions of market movements in a virtual trading environment.

iv. Teaching

20 hours of lectures and 10 hours of seminars in the LT.

Week 11 will be spent working on the extended case study.

v. Formative coursework

Weekly marked problem sheets, with solutions discussed in class. Two marked case studies.

vi. Indicative reading

Lai, T.L. And Xing H. (2008) Statistical Models and Methods for Financial Markets. Springer.

Tsay, R. S. (2005) Analysis of Financial Time Series. Wiley.

Ruppert, D. (2004) Statistics and Finance – an introduction. Springer. Fan, Yao (2003) Nonlinear Time Series.

Hastie, Tibshirani, Friedman (2009) The Elements of Statistical Learning.

Haerdle, Simar (2007) Applied Multivariate Statistical Analysis.

vii. Assessment

Exam (100%, duration: 2 hours) in the summer exam period.

l. Statistical Methods for Risk Management

i. Availability

This course is compulsory on the MSc in Quantitative Methods for Risk Management. This course is available on the Global MSc in Management, Global MSc in Management (CEMS MiM), Global MSc in Management (MBA Exchange), MSc in Data Science, MSc in Financial Mathematics, MSc in Statistics (Financial Statistics) and MSc in Statistics (Financial Statistics) (Research). This course is available as an outside option to students on other programmes where regulations permit.

ii. Pre-requisites

Students must have completed Probability, Distribution Theory and Inference (ST202) and Stochastic Processes (ST302).

ST202, ST302, or equivalent

iii. Course content

A self-contained introduction to probabilistic and statistical methods in risk management. This course starts with risk factors models and loss distributions, which are illustrated via examples in stocks, derivatives, and bonds portfolios. Axioms of coherent risk measures are introduced. Value at risk and other risk measures are introduced and their relation with coherent risk measures is discussed. Multivariate factor models are introduced and analysed: covariance and correlation estimations, multivariate normal distributions and their testing, normal mixture distributions and their fitting to data. The theory of copulas is introduced: meta distributions, tail dependence, fitting copulas to data. Some limitations of copulas are also discussed. The extreme value theory is introduced: generalized extreme value distribution, threshold exceedances and generalized Pareto distribution, modelling and measures of tail risk. Applications to insurance with large loss are also discussed. Students will be exposed to financial data via sets of computer-based classes and exercises.

iv. Teaching

20 hours of lectures and 10 hours of seminars in the MT.

A exercise/problem-solving session will take place in Week 6.

v. Formative coursework

A set of exercises which are similar to problems appearing in the exam will be assigned. A set of coding exercises which are similar to examples in computer lab sessions will be assigned.

vi. Indicative reading

A.McNeil, R.Frey, P.Embrechts, Quantitative Risk Management: Concepts, Techniques, Tools; Princeton Series in Finance

vii. Assessment

Exam (75%, duration: 2 hours) in the January exam period.
Project (25%, 2000 words).

m. Special Topics in Quantitative Analysis: Quantitative Text Analysis

i. Availability

This course is available on the MSc in Applied Social Data Science, MSc in Data Science, MSc in Human Geography and Urban Studies (Research), MSc in Political Science and Political Economy and MSc in Social Research Methods. This course is available with permission as an outside option to students on other programmes where regulations permit.

The course is also available to research students as MY559.

ii. Pre-requisites

Students must have completed Applied Regression Analysis (MY452).

iii. Course content

The course surveys methods for systematically extracting quantitative information from text for social scientific purposes, starting with classical content analysis and dictionary-based methods, to classification methods, and state-of-the-art scaling methods and topic models for estimating quantities from text using statistical techniques. The course lays a theoretical foundation for text analysis but mainly takes a very practical and applied approach, so that students learn how to apply these methods in actual research. The common focus across all methods is that they can be reduced to a three-step process: first, identifying texts and units of texts for analysis; second, extracting from the texts quantitatively measured features - such as coded content categories, word counts, word types, dictionary counts, or parts of speech - and converting these into a quantitative matrix; and third, using quantitative or statistical methods to analyse this matrix in order to generate inferences about the texts or their authors. The course systematically surveys these methods in a logical progression, with a practical, hands-on approach where each technique will be applied using appropriate software to real texts.

Lectures, class exercises and homework will be based on the use of the R statistical software package but will assume no background knowledge of that language.

iv. Teaching

20 hours of lectures and 10 hours of computer workshops in the LT.

v. Formative coursework

Exercises from the computer classes can be submitted for marking.

vi. Indicative reading

quanteda: An R package for quantitative text analysis. http://kbenoit.github.io/quanteda/

Grimmer, Justin and Brandon M. Stewart. 2013. “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.” Political Analysis 21(3):267–297.

Loughran, Tim and Bill McDonald. 2011. “When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks.” The Journal of Finance 66(1, February): 35–65.

Evans, Michael, Wayne McIntosh, Jimmy Lin and Cynthia Cates. 2007. “Recounting the Courts? Applying Automated Content Analysis to Enhance Empirical Legal Research.” Journal of Empirical Legal Studies 4(4, December):1007–1039.

vii. Assessment

Project (40%, 3000 words) and coursework (60%, 2000 words) in the ST.

n. Algorithms and Computation

i. Availability

This course is available on the MSc in Applicable Mathematics, MSc in Data Science, MSc in Operations Research & Analytics, MSc in Statistics, MSc in Statistics (Financial Statistics), MSc in Statistics (Financial Statistics) (Research) and MSc in Statistics (Research). This course is available with permission as an outside option to students on other programmes where regulations permit.

The course is compulsory for students on the MSc Applicable Mathematics who are not taking MA421 Advanced Algorithms; it is optional for students on the MSc Applicable Mathematics who take the advanced course.

ii. Pre-requisites

Good general knowledge of mathematics, including familiarity with abstract concepts. A willingness to cope with technical details of computer usage, and with a rapid introduction to programming.

iii. Course content

Introduction to programming in Java. Introduction to the theory of algorithms: running time and correctness of an algorithm. Recursion. Data structures: arrays, linked lists, stacks, queues, binary search trees. Sorting algorithms. Greedy algorithms. Dynamic programming. Inheritance and Generics in Java.

iv. Teaching

20 hours of lectures, 20 hours of seminars, 9 hours of workshops and 10 hours of help sessions in the MT. 2 hours of lectures in the ST.

Workshops will be held before the start of MT.

v. Formative coursework

Weekly exercises are set and marked. Many of these will require implementation of programming exercises in Java.

vi. Indicative reading

T H Cormen, C E Leiserson, R L Rivest and C Stein, Introduction to Algorithms;

R Sedgewick, K Wayne, Introduction to programming in Java; D Flanagan, Java in a Nutshell.

vii. Assessment

Exam (75%, duration: 2 hours and 30 minutes) in the summer exam period. Coursework (25%) in the MT.

o. Modelling in Operations Research

i. Availability

This course is compulsory on the MSc in Management Science (Decision Sciences) and MSc in Operations Research & Analytics. This course is available on the Global MSc in Management, Global MSc in Management (CEMS MiM), Global MSc in Management (MBA Exchange) and MSc in Data Science. This course is available with permission as an outside option to students on other programmes where regulations permit.

ii. Pre-requisites

Students must know basics of linear algebra (matrix multiplication, geometric interpretation of vectors) and probability theory (expected value, conditional probability, independence of random events).

Students taking the course as an outside option are also expected to have a basic knowledge of linear programming. For students in the MSc in Operations Research & Analytics this will be covered in MA423 Fundamentals of Operations Research.

iii. Course content

The course will be in 2 parts, covering the two most prominent tools in operational research: simulation, the playing-out of real-life scenarios in a (computer-based) modelling environment, and mathematical optimisation, the application of sophisticated mathematical methods to make optimal decisions.

Simulation (8 lecture hours): This part develops simulation modelling skills, understanding of the theoretical basis which underpins the simulation methodology, and an appreciation of practical issues in managing a simulation modelling project. Topics include Monte Carlo simulation, Markov processes, discrete event simulation, and variance reduction. The course will teach students how to use a simulation modelling software package.

Optimisation (12 lecture hours): This part enables students to model and solve real-life management problems as Mathematical Optimisation problems. In providing an overview of the most relevant techniques of the field, it teaches a range of approaches to building Mathematical Optimisation models and shows how to solve them and analyse their solutions. Content includes: The modelling life-cycle and modelling environments; formulation of management problems using linear and network models; solution of such problems with a special-purpose programming language; interpretation of the solutions; limitations of such models; and formulation and solution of nonlinear models including some or all of binary, integer, convex and stochastic programming models.

v. Teaching

20 hours of lectures, 13 hours and 30 minutes of seminars and 10 hours of computer workshops in the MT. 8 hours of computer workshops in the LT.

Computer workshops are not mandatory. They are help sessions, where an instructor is available to students in the computer cluster while they work on their assignment.

vi. Formative coursework

Students will be expected to produce 1 project in the MT.

Feedback will be provided on the weekly homework. Additional feedback will be provided on a one-on-one basis to students attending the optional computer help sessions.

vii. Indicative reading

Full lecture notes will be provided to students for both parts.

Recommended readings:

Simulation

A M Law & W D Kelton, Simulation Modelling and Analysis, McGraw Hill (3rd ed., 2000);

M Pidd, Computer Simulation in Management Science, Wiley (5th ed., 2006);

S Ross, Simulation, Academic Press (5th ed., 2012).

Optimisation

W L Winston, Operations Research: Applications and Algorithms, Brooks/Cole (4th ed., 1998);

D Bertsimas and J N Tsitsiklis: Introduction to Linear Optimization, Athena Scientific (3rd ed., 1997).

viii. Assessment

Project (100%) in the LT.

The project will be on Simulation, Mathematical Optimisation, or a combination of the two. The deliverable is a report of at most 12 pages (main report, excluding executive summary and technical appendices), along with a soft copy of any computer code and solver output.