M2 ECONOMÉTRIE DATA SCIENCE (EDS)
M2 ECONOMETRICS AND DATA SCIENCE (EDS)

Cours spécifiques à l'alternance

Cours spécifiques à l'option Magistère Economie, Data Science et Finance

Dernière mise à jour / last update: septembre 2024.

Machine Learning and Statistical Learning
Machine learning et statistical learning

COURSE LANGUAGE
English

TEACHER
Pierre MICHEL – pierre.michel@univ-amu.fr

COURSE DESCRIPTION AND OBJECTIVES
This course introduces the fundamental concepts of machine learning and statistical learning. Its primary goal is to facilitate a comprehensive understanding of how machine learning algorithms operate. Students will explore a variety of widely used methods for supervised tasks. The course emphasizes practical implementation by providing hands-on experience with both synthetic and real-world datasets.

COURSE OUTLINE
1. Introduction
2. Estimation of the Parameters: an Optimization Problem
3. Regression Tasks
4. Classification Tasks
5. Explainable Machine Learning
6. Machine Learning, Ethics, and Fairness

KEY PROFESSIONAL SKILLS UPON GRADUATION
Understanding how machine learning algorithms work
Understanding the basic theory underlying machine learning
Being able to code (in Python or R – choice is up to the students) simple machine learning algorithms.

ORGANIZATION
Semester: S1
Teaching Hours: 24 h of lectures, sessions alternating theoretical presentations and applications.
Comment: The applications will be carried out on personal computers.

BIBLIOGRAPHY AND TEXTBOOKS
Berk, R. A. (2016). Statistical Learning from a Regression Perspective. Springer Texts in Statistics. doi:10.1007/978-3-319-44048-4
Charpentier, A. (2024). Insurance, Biases, Discrimination and Fairness. Springer
James, G, Witte, D., Hastie, T., Tibshirani, R. (2023). An Introduction to Statistical Learning with Applications in R. Second Edition. Springer
Murphy, K. (2012). Machine Learning: A Probabilistic Perspective. The MIT Press. ISBN: ISBN: 9780262018029

Coding
Programmations

COURSE LANGUAGE
English

ENSEIGNANT / TEACHER
Flavien RICHE – practitioner

COURSE DESCRIPTION AND OBJECTIVES
The aim of this course is to enable students to build a complete data pipeline from the data source to the user interface. This pipeline is implemented via an initial Web scrapping stage, followed by a transformation stage and finally a stage in which key information is rendered in a dashboard. By the end of the course, students will be able to serve a customer and deliver an end-to-end solution. Each module of the course is associated with a series of exercises to ensure the continuous implementation of the concepts studied.

COURSE OUTLINE
1) Python
   - 1.1 Review of Python basics (objects, lists, dictionaries, etc.)
   - 1.2 Essential libraries (Pandas, Numpy, Sklearn, Beautifulsoup, etc.)
   - 1.3 Focus on webscrapping
   - 1.4 Programming rules and code quality
   - 1.5 Using the Windows/Linux terminal
   - 1.6 Introduction to creating command lines
   - 1.7 Introduction to creating python libraries.
2) Flexdashboard (R)
   - 2.1 Introduction to Flexdashboard
   - 2.2 Building a dashboard with Flexdashboard
   - 2.3 Orchestrating the python pipeline via system commands from R

KEY PROFESSIONAL SKILLS UPON GRADUATION
Creation of libraries, iImproving code quality, use of the terminal dashboarding, webscrapping, intermediate level Python, pipeline automation and orchestration

ORGANIZATION
Semester: S1
Teaching Hours: 24 h of lectures
Comment: Each module is followed by application exercises. A mini project is also included during the course to develop students' analytical skills.
Examination Method: Group project which aims to extract data from the web using webscrapping methods, generalise the extraction by creating a dedicated library, clean the data, create KPIs and display them in a dashboard.

BIBLIOGRAPHY AND TEXTBOOKS
https://docs.python.org/
https://pkgs.rstudio.com/flexdashboard/
https://pypi.org/project/beautifulsoup4/

MANDATORY PREREQUISITES
Python for beginner, R for beginner, Intermediate level Jupyter Notebook, Intermediate level of VS Code

RECOMMENDED PREREQUISITES
Basic knowledge of terminal utilisation

KEYWORDS
Python, dashboard, webscrapping, terminal, libraries

Softwares
Logiciels

COURSE LANGUAGE

English

TEACHER

Virgile PESCE – practitioner

COURSE DESCRIPTION AND OBJECTIVES

The objective of this course is to teach Python and its most used libraries for data analysis and data visualization. It aims to make students masters in data handling, manipulation, and visualization.

COURSE OUTLINE

Basics of web: HTML, CSS, Bootstrap, deployment
Web scraping using Python and its limits
Data handling using Pandas
Data visualization using Python, best practices
Building interactive and dynamic dashboards using Dash from Plotly
Git, GitHub, GitLab for code management
Data Analytics Softwares

ORGANIZATION

Semester: S1
Teaching Hours: 24 h of lectures

Dernière mise à jour / last update: juin 2025.

Predictive methods
Méthodes de prévisions

COURSE LANGUAGE

English

TEACHER

Christophe MULLER – christophe.muller@univ-amu.fr

COURSE DESCRIPTION AND OBJECTIVES

The main objective of this course is to provide the students with a synthetic framework so that they can theoretically understand and efficiently apply the main forecast and prediction techniques for most applied econometric problems. An important aim is to make students able to think by themselves when facing an applied econometric problem to solve.

PLAN DU COURS / COURSE OUTLINE

I. Introduction to Economic Predictions
- Uncertainty and Information
- Cross Sections, Time Series and Panel Data
- Notions of Prediction and Forecast
- Reminders of probability, algebra and optimisation

II. Expectations and Regressions
- Reminders of linear regression
- Consistency and Asymptotic Normality
- Simulators

III. Predictions
- Statistical decision theory
- Linear Predictors
- Evaluating Predictive Power
- Identification
- Nonlinear predictions: example of the nonlinear Tobit model

IV. Dynamic Forecasts without Time Series Models
- Sequences
- Time operators, moving averages and autocorrelations
- Extrapolation and smoothing
- Seasonality correction and trends

V. Dynamic Forecasts with Univariate Time Series Models
- Forecasts with stationary ARIMA models
- Evaluating forecasts

VI. Dynamic Forecasts with Multivariate Time Series Models
- VAR models
- Forecasting under non-stationarity
- Panel data

VII. Poll Surveys and Experts
- Inference for statistical surveys based on random samples
- Quotas and redressing
- Expert advice
- Structural models

VIII. Predictions with Big Data and Machine Learning
- Machine learning for predictions
- Variable selection and LASSO
- Neural networks
- Artificial intelligence

IX. Empirical Applications
- Empirical examples
- The command forecast in Stata
- The forecasting package in R
- Poll internet survey
- Treatment effect and external validity
- Volatility forecast

- Use of prediction statistics by stakeholders

KEY PROFESSIONAL SKILLS UPON GRADUATION

The targeted skills are an understanding of the reasonings made when using econometric techniques. At the end of the course, the students should be able to make their own informed methodological decisions in empirical applications, notably for prediction and forecasts.

ORGANIZATION

Semester: S1
Teaching Hours: 24 h of lectures
Examination Method: Report of prediction analysis

BIBLIOGRAPHY AND TEXTBOOKS

Elliott, G. and A. Timmerman, Economic Forecasting, Journal of Economic Literature, February 2008.
Ghysels, E. and M. Marcellino, Applied Economic Forecasting Using Time Series Methods, Oxford University Press, 2018.
Gouriéroux, C. and A. Monfort, Time Series Econometrics, Cambridge University Press, 1996.
Hyndman, R.J., Forecasting: Principles and Practice, 2024.
Hamilton, J.D., Time Series Analysis, Princeton University Press, 1994.
Montgomery, D.C., M. Kulahci and C.L. Jennings, Time Series Forecasting, 2024IIU
Stata Manual, 2024
Wooldridge, J.W., Introduction Econometrics: A Modern Approach, South-Western College Publishing, 2012.

MANDATORY PREREQUISITES

Linear econometrics, basic mathematics of the optimisation, linear algebra, Stata software, basic time series econometrics.

KEYWORDS

Econometrics, Identification, Inference, Estimation & Prediction, Forecasts, Econometric models, Big Data.

Automatic Model Selection Methods
Modèles de réduction de l'information

COURSE LANGUAGE

English

TEACHER

Sullivan HUE – sullivan.hue@univ-amu.fr

COURSE DESCRIPTION AND OBJECTIVES

The objective of this course is to introduce quantitative methods allowing to reduce information. These methods cover different fields of statistics and are based on classical econometric methods (OLS, MLE) or classificatory or principal component methods. The goal is to study methods to do automatic variable selection in large-scale problems and to apply them to real data.

COURSE OUTLINE

- Classification methods
- Statistical factor models
- Lasso methods
- The so-called « General to Specific » method (Hendry, Gets or Autometrics Methodology)

KEY PROFESSIONAL SKILLS UPON GRADUATION

Understanding new methods
Application on real data
Learning new tools or econometric softwares dedicated to the reduction of information

ORGANIZATION

Semester: S1
Teaching Hours: 24 h of lectures

BIBLIOGRAPHY AND TEXTBOOKS

Doornik, J.A. and Hendry, D.F. (2015). Statistical model selection with "Big Data", Cogent Economics & Finance, vol 3, n°1, 1-15.
Hendry, D.F. and Doornik, J.A. (2014). Empirical Model Discovery and Theory Evaluation. Automatic Selection Methods in Econometrics. The MIT Press.
Richard A. Johnson and Dean W. Wichern, Applied Multivariate Statistical Analysis, Pearson.

Non-Parametric Methods in Econometrics
Méthodes non paramétriques en économétrie

COURSE LANGUAGE

English

TEACHER

Costin PROTOPOPESCU – costin.protopopescu@univ-amu.fr

COURSE DESCRIPTION AND OBJECTIVES

Non-Parametric methods are statistical techniques that do not require to specify functional forms for objects being estimated. Instead, they let the data itself plays and informs the resulting model in a particular manner. Such methods are becoming increasingly popular for applied data analysis, they are best suited to situations involving large data sets for which the number of variables involved is manageable. These methods are often deployed after common parametric specifications are found to be unsuitable for the problem at hand, particularly when formal rejection of a parametric model based on specification tests yields no clues as to the direction in which to search for an improved parametric model.
The job market understood the importance of the non/semi-parametric methods and almost any serious software contains the principal techniques in this area. We illustrate the different models and techniques with R and Matlab. First, R because of the huge number of packages from CRAN, and secondly Matlab because is the easiest environment for programming arrays in econometrics (and typically all objects are arrays in applied econometrics). Both are very representative for the job market. Each lecture will be accompanied by numerical examples and small programming tutorials.

KEY PROFESSIONAL SKILLS UPON GRADUATION

To master the concepts specific to nonparametric methods for modelling the conditional expectation and the conditional variance.
Estimation and forecasting in i.i.d and non i.i.d. simulated or real data sets.
Be able to use the nonparametric estimation for driving more structural models.
To use the basic packages from R and the toolboxes from Matlab for nonparametric and semiparametric modelling.

ORGANIZATION

Semester: S1
Teaching Hours: 24 h of lectures

BIBLIOGRAPHY AND TEXTBOOKS

Given the wide content of the course, there is no single textbook for this course, therefore the lecture handouts are self-contained. However, the content is covered in some classic textbooks, such as:
Jeff RACINE: “Nonparametric Econometrics, A Primer”, Foundations and Trends in Econometrics, Vol. 3, No 1 (2008).
Adrian PAGAN and Aman ULLAH, “Nonparametric Econometrics”, Cambrdige University Press, (1999).
B. P. SILVERMAN: “Density Estimation for Statistics and Data Analysis”, Chapman Hall, (1986).
Wolfgang HÄRDLE: “Applied Nonparametric Regression”, Cambridge University Press; Revised edition (1992).
Denis BOSQ: “Nonparametric Statistics for Stochastic Processes: Estimation And Prediction”, Springer-Verlag New York Inc, 2nd ed. (2013).
Ziyue LIU and Wensheng GUO: “Data Driven Adaptive Spline Smoothing”, Statistica Sinica 20 (2010).

Dernière mise à jour / last update: juin 2025.

Time Series
Séries temporelles

COURSE LANGUAGE

English

TEACHER

Gilles DUFRENOT – gilles.dufrenot@univ-amu.fr

COURSE DESCRIPTION AND OBJECTIVES

The objective of this course is to provide the students with some tools that are necessary for those who will be confronted in their future careers with functions requiring knowledge in quantitative finance and quantitative economics analysis. The course includes theoretical and applied aspects using R software.

COURSE OUTLINE

1.- Extreme distributions and copula analysis
2.- Regime-switching non-linear models: Markov-switching and STAR models
3.- Long-memory models

KEY PROFESSIONAL SKILLS UPON GRADUATION

Acquire complex tools for analyzing time series
Learn how to use the R programming language to estimate nonlinear models
Learn how to identify unusual behavior in series: regime shifts, hysteresis, extreme events.

ORGANIZATION

Semester: S1
Teaching Hours: 24
Examination Method: Final Exam + Work in group

BIBLIOGRAPHY AND TEXTBOOKS

G. Dufrénot and T. Matsuki (eds), 2021, Recent Econometric Techniques for Macroeconomics and Financial Data, Springer Verlag (see the chapter written by Aditi Chaubal).
Elements of copula Modelling with R (https://copula.r-forge.r-project.org/book/)
P. Robinson (ed), 2023, Time series with long-memory, Oxford University Press (https://academic.oup.com/book/51958)

MANDATORY PREREQUISITES

Students must have basic knowledge of time series: dynamic modelling, ARIMA models, stochastic and deterministic stationarity. In addition, knowledge of programming in R is required. Familiarity with elementary probability distributions and laws is also essential.

RECOMMENDED PREREQUISITES

Knowledge of spectral analysis is recommended for analysing time series in the frequency domain (spectral density, power spectrum, spectral windows).

KEYWORDS

Markov-switching, Copula, ARFIMA, Long-Memory, nonlinearity

Dernière mise à jour / last update: juin 2025.

Practising Data Science in the Real World: Limitations and Challenges (non MAG)
Utilisation de la science des données : limites et défis (non MAG)

COURSE LANGUAGE

English

TEACHER

Dimitri SCRONIAS – practitioner from INSERM

COURSE DESCRIPTION AND OBJECTIVES

The use of data science techniques faces several limitations and challenges that require a multidisciplinary approach, incorporating technical expertise, domain knowledge, ethical considerations, and effective communication and collaboration skills. By acknowledging and mitigating these challenges, data scientists can maximize the value and impact of data science in real-world applications while ensuring responsible and ethical
use of data and technology.

COURSE OUTLINE

Class 1: How to prepare a data science project? Defining a business problem, preparing a data analysis plan, and overview of the methodologies and tools at the disposal of the data scientist.
Lab 1: Refresher on R and the tidyverse packages.
Class 2: Cleaning "bad data". Definitions of data cleaning, types of "bad data" (missing data, outliers, measurement errors/biased data, non-machine-readable data…) and how to mitigate or solve these issues.
Lab 2: Data cleaning and imputation on a messy dataset to aggregate from different sources, with various problems and multiple patterns of missing data and biases.
Class 3: Focus on different types of bias, bias in the context of causal inference, and methods to solve potential bias issues.
Fairness in Machine Learning: what is fairness, why it matters, and metrics to measure it.
Class 4: Good practices on communicating results to decision-makers and making impactful data visualisations (tables, graphs, dashboards or regression/ML model results).
Presentation of model-agnostic interpretation tools
Lab 3: Interactive dashboard from scratch with R, using the *shiny* and *bslib* packages.
Class 5/Lab 4: Methods and tools to improve your programming skills, the performance of your code, and to handle larger-than-memory datasets.

KEY PROFESSIONAL SKILLS UPON GRADUATION

Being able to develop critical thinking about the use of data science techniques.
Being able to approach problems systematically and identify alternative solutions
Understanding ethical considerations and implications in data science, such as privacy, fairness, and transparency
Knowing how to communicate insights, limitations, and recommendations clearly

ORGANIZATION

Semester: S1
Teaching Hours: 24 h of lectures, 8 sessions of 3 hours each
Examination Method: Partial continuous assessment

BIBLIOGRAPHY AND TEXTBOOKS

Foster Provost and Tom Fawcett, Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking, 2013
Emmanuel Ameisen, Building Machine Learning Powered Applications: Going from Idea to Product, O'Reilly Media, 2020
Q. Ethan McCallum, Bad Data Handbook, O'Reilly Media, 2012
Herbert Weisberg, Bias and Causation: Models and Judgement for Valid Comparisons, Wiley, 2010
Scott Cunningham, Causal Inference: The Mixtape, Yale University Press, 2021
Solon Barocas, Moritz Hard and Arvind Narayanan, Fairness and Machine Learning: Limitations and Opportunities, MIT Press, 2023
Claus O. Wilke, Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures, O'Reilly Media, 2019
Cole Nussbaumer Knaflic, Storytelling with Data: A Data Visualization Guide
Christoph Molnar, Interpretable Machine Learning: A guide For Making Black Box Models Explainable, 2022

Hadley Wickham, Advanced R, Chapman & Hall/CRC Press, 2019
Robert Cecil Martin, Clean Code: A Handbook of Agile Software Craftsmanship, Prentice Hall, 2008

RECOMMENDED PREREQUISITES

Basic R proficiency (Tidyverse is a plus)

KEYWORDS

Data science; bias; data cleaning; communication; R

Application: Quantitative Marketing (non MAG)
Application : Marketing quantitatif (non MAG)

COURSE LANGUAGE
English

TEACHER
Benoît HUBERT – practitioner from IPSOS
Yaroslav KOZYREV – practitioner from IPSOS
Bora NUMANI – practitioner from IPSOS

COURSE DESCRIPTION AND OBJECTIVES
This course provides M2 students introduction with knowledge and applied skills to design, develop, and deploy agentic systems that leverage big data for marketing purposes. Throughout the sessions, students will be offered the opportunity to realize an agentic project. Within this framework, they will develop specific agents tailored to marketing challenges and explore the fundamental techniques related to the deployment of agentic solutions, including but not limited to Agent-to-Agent (A2A) architectures and Multi-Component Pipelines (MCP). The course places a strong emphasis on the practical integration of recent innovations in agentic system design and the application of these solutions to real-world marketing use cases. The pedagogical approach combines theoretical lectures, critical analysis of case studies, and hands-on group projects, with the objective of equipping students to become proficient in the design, orchestration, and management of agentic Marketing AI systems.

COURSE OUTLINE
The course begins with an introduction to Big Data in Marketing, presenting the intersection of modern machine learning techniques and their applications to marketing problems, alongside a discussion of ethical considerations. The following sessions examine the conceptual frameworks for agentic marketing solutions, including the analysis of successful case studies and the assessment of their marketing impact. Students are then introduced to the data processing and automation tools required for the implementation of agentic systems, with a focus on Python, agent frameworks, cloud providers, and database technologies. A dedicated module explores the design, orchestration, and deployment of agentic systems, with particular attention to advanced methods such as Agent-to-Agent (A2A) communication and Multi-Component Pipelines (MCP). Students will be guided through the process of building and scaling their own agentic projects, integrating AI components and deploying them in realistic marketing contexts. The course concludes with the demonstration and critical evaluation of group project prototypes, with an emphasis on innovation, robustness, and marketing impact.

KEY PROFESSIONAL SKILLS UPON GRADUATION
Upon completion of the course, students will have acquired practical experience in the development and deployment of agentic systems for marketing, proficiency in handling Large Language Models (LLMs) and Natural Language Processing (NLP) for automation and decision-making, advanced skills in product and project management within technological environments, expertise in the use of cloud and database tools, and the ability to apply state-of-the-art deployment strategies such as Agent-to-Agent (A2A) communication and Multi-Component Pipelines (MCP) for scalable, robust marketing solutions.

ORGANIZATION
Semester: S1
Teaching Hours: 24 h of lectures
Examination Method: Presentation of a group project

MANDATORY PREREQUISITES
A working knowledge of Python, a basic understanding of Large Language Models, APIs, and a willingness to engage with new technologies are required.

RECOMMENDED PREREQUISITES
NLP, LLMs, Agentic Systems, Cloud Programming, data management

KEYWORDS
Marketing, Big Data, Machine Learning, Large Language Models, Agents, Cloud.

Methodology of Econometric and Statistical Studies
Méthodologie des études économétriques et statistiques

COURSE LANGUAGE

English

TEACHER

Nathalie UDUMYAN-MESROBIAN – practitioner

COURSE DESCRIPTION AND OBJECTIVES

This course aims to provide a structured methodology for designing and carrying out applied econometric and statistical studies. The course presents the different methods of data collection, analysis and interpretation, while addressing the challenges encountered in a professional setting (regulatory and ethical concerns, project management, communication of results). This course aims to enable students to master responding to a call for tenders to carry out an econometric or statistical study, from defining the topic and terms of reference to preparing a technical and financial proposal.

COURSE OUTLINE

1. Introduction
   - Call for tenders: definition, process and examples
2. Group work session: developing terms of reference
3. Design of an econometric/statistical study
4. Data collection methods, ethical and regulatory concerns
5. Group work session: methodological proposal
6. Organizational aspects
   - Project management and preparation of financial proposal
   - Communication of results

KEY PROFESSIONAL SKILLS UPON GRADUATION

To develop a comprehensive methodology for data collection, processing and analysis.
To identify and apply the most appropriate statistical or econometric methods to a given problem.
To master the methodological stages of an applied econometric or statistical study.
To develop terms of reference and respond to a call for tenders on a topic of interest, including the drafting of a methodological and financial proposal.
To communicate effectively on methods and results to various audiences, specialists and non-specialists alike, in a professional context.
To work as part of a team to design and present a structured project.

ORGANIZATION

Semester: S1
Teaching Hours: 24 h of lectures, 6 sessions of 4 h each.
Comment: Alternating lectures and supervised group work to produce and present the terms of reference for a call for tenders and the response to this call.
Examination Method: Continuous assessment: terms of reference (50%) and oral presentation of proposed methodology (50%) + Final assessment: methodological proposal (50%) and defence in front of a jury (50%).

BIBLIOGRAPHY AND TEXTBOOKS

Material provided by teachers.

MANDATORY PREREQUISITES

Mastery of econometric and statistical methods and ability to work in English

RECOMMENDED PREREQUISITES

Interest in applied economic issues

KEYWORDS

Applied econometrics and statistics, methodology of empirical studies, Terms of reference, project management

Professionalisation Workshops
Ateliers de professionnalisation

COURSE LANGUAGE

English

COURSE DESCRIPTION AND OBJECTIVES

This workshop is designed to guide students in their transition from academic training to the job market. **Participation in all activities is mandatory.**

It combines several complementary components:

- **Afterworks** (on campus or online), where companies and institutions introduce themselves to students, share insights into their missions, and discuss opportunities for collaboration.
- A **Career Day**, organized in two parts: first, recent graduates present their career paths, current positions, and how their training helped them enter the job market; second, a large recruitment fair brings together around 50 local, national, and international companies and institutions to offer internships and job opportunities.
- A course entitled *"Building a Strong Application"*, providing practical tools and strategies for professional integration. It is divided into two parts:
  - First part (lecture): Preparing for interviews (best practices, preparation methods, and self-presentation); searching for an internship or a job abroad (application strategies, networks, and resources); negotiating salaries (key principles for successful negotiation).
  - Second part (workshops): Small-group sessions offered to M2 students, focusing on CV writing and mock interview practice.

Together, these activities give students concrete experience, direct contact with employers, and essential skills to confidently approach their future careers.

KEY PROFESSIONAL SKILLS UPON GRADUATION

By the end of the workshop, students will possess the essential skills to approach the job market with confidence. They will know how to present themselves effectively, understand recruiters' expectations in France and abroad, and activate a professional network. Through lectures and practical workshops, they will be able to prepare strong applications, succeed in interviews, and conduct salary negotiations with assurance.

ORGANIZATION

Semester: S1
Teaching Hours: 10 h of tutorials

Advanced Machine Learning
Machine learning avancé

COURSE LANGUAGE

English

TEACHER

Badih GHATTAS – badih.ghattas@univ-amu.fr

COURSE OUTLINE

Introduction to Machine Learning principles, Supervised and Unsupervised
Regression and classification Trees
Ensemble methods, Bagging boosting and random Forests
Kernel approaches for linear separation
Neural Networks and uses cases.

KEY PROFESSIONAL SKILLS UPON GRADUATION

Implementing a Machine Learning or Deep Learning approach with any dataset to resolve a practical problem after clearly identifying its context and type.

ORGANIZATION

Semestre : S2
Charge d'enseignement : 24 h de cours magistraux
Examination Method: Oral presentation of a project treated by groups of a maximum of three students during all the lab classes.

MANDATORY PREREQUISITES

Programming with either R or Python

Dernière mise à jour / last update: juin 2025.

Interpretability and Causality in Machine Learning
Interprétabilité et causalité en machine learning

COURSE LANGUAGE

English

TEACHER

Emmanuel FLACHAIRE – emmanuel.flachaire@univ-amu.fr
Sullivan HUE – sullivan.hue@univ-amu.fr

COURSE DESCRIPTION AND OBJECTIVES

This course is divided into two parts: (1) understanding why machine learning algorithms are not interpretable and how to use interpretable methods to interpret black boxes; (2) introducing recent advances in causal machine learning.

COURSE OUTLINE

1. Interpretability
   - Interpretable models
   - Feature effects methods: PDP, ICE and ALE
   - Surrogate models
   - Shapley Values based methods
   - IML limits
   - Inherently interpretable models with high performance
2. Causality
   - Directed Acyclic Graphs (DAG) and potential outcomes approaches
   - Regression and double orthogonalization
   - IPW and AIPW estimators
   - Post-Lasso and Post-double Lasso
   - Double/debiased machine learning
   - Generic machine learning for CATE

ORGANIZATION

Semester: S2
Teaching Hours: 24 h of lectures

BIBLIOGRAPHY AND TEXTBOOKS

Molnar (2025) Interpretable machine learning
Chernozhukov et al (2024) Causal ML book
Facure (2023) Causal inference for the brave and true
GHaillac and L'Hour (2025) Machine learning for econometrics

Transition and Duration Models
Modèles de transitions et de durées

COURSE LANGUAGE
English

TEACHER
Christian SCHLUTER – christian.schluter@univ-amu.fr

COURSE DESCRIPTION AND OBJECTIVES
Students will study models of transitions and durations and learn how to estimate these using real-world data.
This course is an introduction to modelling transitions into a state of interest (such as the transition into employment from unemployment) and durations (such as unemployment, survival of patients after medical treatment or firms after a financial crash, time-to-default of loans, or time-to-purchase, criminal recidivism). Time-to-event or survival analysis are alternative labels. We start with the basic building blocks (Poisson processes, Markovian transitions, Markov chains, hazard models). Since duration data might be censored (individuals might still be in the state of interest at the end of the observation window), classic ordinary least squares (OLS) is invalid, and we develop appropriate methods for estimation. Unobserved heterogeneity introduces fundamental identification challenges (duration dependence v. dynamic sorting) that are discussed in detail. Finally, we consider how recent machine learning methods have been adapted for such censored duration data (such as Random Survival Forests).
Throughout all methods will be illustrated using examples in R and python, and we will replicate several papers from the established empirical literature. Several exercise sets will help students deepen their understanding of the theory.

COURSE OUTLINE
(I) Introduction to Poisson and counting processes
- Counting processes and the Poisson process
- Exponentially distributed inter-arrival times / durations
- The Poisson process and Markov chains
- Poisson regressions
- The Piece-Wise Exponential (PWE) model and estimation using a GLM

(II) Introduction to Markov processes
- Transitions, and the Chapman-Kolmogorov equation
- Classification of states and first passage or hitting times
- The invariant distribution
    - Markov's (or the ergodic) theorem
- Examples in theory and practice
    - State-space modelling: MC approximation to an AR(1) process (Rouwenhorst method)
    - State-space modelling using the Poisson process.
    - Unemployment transitions
    - Google's PageRank

(III) Duration and survival analysis: Hazard models
- Survival functions: The Kaplan-Meier estimator, the log-rank test
- Hazards, and the Proportional Hazard (PH) model
    - Maximum likelihood estimation (flow and stock samples)
- The Mixed Proportional Hazard (MPH) model, identification challenges
- The PH model and grouped data
- Cox's Partial Likelihood
- Machine Learning and Survival Analysis
    - Training a PH model
    - Random Survival Forests

KEY PROFESSIONAL SKILLS UPON GRADUATION
The students will master the theory of transition and duration models, learn how to estimate such models in practice using real-world data, and understand and address the empirical challenges that arise in empirical work.

ORGANIZATION
Semester: S2
Teaching Hours: 24 h of lectures
Examination Method: Research Project + Exam.

BIBLIOGRAPHY AND TEXTBOOKS

G. James et al. (2021): An Introduction to Statistical Learning, Chapter 11.
Wooldridge (2002), Chapters 19 and 20,
van den Berg, Chapter 55, Handbook of Econometrics.

MANDATORY PREREQUISITES

Basic econometrics.

KEYWORDS

Transition models, duration models, Poisson process, hazard and survival functions.

Hackathon and Certification

COURSE LANGUAGE

English

COURSE DESCRIPTION AND OBJECTIVES

This is an independent non-supervised activity, where students participate (typically in small groups) in a public "Hackathon", i.e. "an event, typically lasting several days, in which a large number of people meet to engage in collaborative computer programming.” Competition themes and organisers can change from year to year.
Examples from the past: In the last few years Airbus has invited M2 AMSE students to compete against other universities in hackathons focusing on Computer Vision and Natural Language Processing.

"Certification" is an independent non-supervised activity where students seek to obtain a certification as "professional" data scientists. Providers of such a certification can change from year to year.
This only concerns non MAG classical track students.

Professional Communication (apprenticeship)
Communication professionnelle (en alternance)

COURSE LANGUAGE

English

TEACHER

Julia MALEK – practitioner

COURSE DESCRIPTION AND OBJECTIVES

This course is designed to help students build a confident, strategic and authentic professional presence. Through an engaging, practice-oriented approach, students will sharpen their communication skills for key professional contexts: job interviews, personal branding, salary negotiations, public speaking, and professional networking (both in-person and online). At the end of the course, each student will have crafted their own impactful professional narrative and acquired concrete tools to thrive in their future career.

COURSE OUTLINE

Pillar 1 — Self-knowledge & Personal Branding Foundations
Pillar 2 — Pitching Yourself Professionally
Pillar 3 — Job Interviews & Recruitment Processes
Pillar 4 — Professional Digital Presence & LinkedIn
Pillar 5 — Ethical Negotiation Strategies
Pillar 6 — Nonviolent Communication in the Workplace

KEY PROFESSIONAL SKILLS UPON GRADUATION

Communicate clearly and persuasively in professional settings
Build a coherent and compelling personal brand, online and offline
Prepare for and manage job interviews with confidence, including tricky and unexpected questions
Approach salary negotiations with strategy, ethics and assertiveness
Pitch themselves with clarity and impact (voice, structure, body language)
Better understand their communication and decision-making style using MBTI insights
Integrate Nonviolent Communication (NVC) techniques to build respectful, constructive dialogue
Reinforce written communication tools (CV, cover letter) through spoken delivery and rhetorical techniques

ORGANIZATION

Semester: S2
Teaching Hours: 20 h of tutorial sessions
Comment: For students doing an apprenticeship. Interactive workshops combining theory, active exercises, and coaching

BIBLIOGRAPHY AND TEXTBOOKS

Les mots sont des fenêtres (ou bien ce sont des murs), Marshall B. Rosenberg

Scientific Communication (apprenticeship)
Communication scientifique (en alternance)

COURSE LANGUAGE
English

TEACHER
Morgan RAUX – morgan.raux@univ-amu.fr

COURSE DESCRIPTION AND OBJECTIVES
This course provides graduate students with the skills necessary to effectively communicate scientific ideas to diverse audiences, including academics, policymakers, multi-disciplinary audiences, and the general public. Teaching modules will use a range of discursive (focused on theoretical issues) and practical (focused on skills) approaches.

COURSE OUTLINE
Possible modules include: Foundations of Scientific Communication; Writing for Scientific Audiences; Visual and Oral Communication; Communicating with Non-Specialist Audiences.

KEY PROFESSIONAL SKILLS UPON GRADUATION
By the end of this course, students will be able to write clear, concise, and compelling scientific papers/reports and abstracts, and deliver effective oral presentations. They will also be able to write a review/discuss someone else's work and respond to a reviewer. We will pay particular attention to the audience to which a student will communicate. Students will be able to navigate around some digital tools designed for scientific communication.

ORGANIZATION
Semester: S2
Teaching Hours: 24 h of tutorial sessions
Comment: For students doing an apprenticeship. This course is designed to be highly interactive, with an emphasis on practical applications. Students are encouraged to bring their own topics and projects into the coursework to make the learning experience as relevant as possible. Active participation in class is mandatory. The modules will be introduced through activities. Assignments will be to a large extent completed during class time, including short writing assignments and oral presentations.
Examination Method: ontinuous assessment

BIBLIOGRAPHY AND TEXTBOOKS
Mercer-Mapstone, Lucy, and Louise Kuchel. "Core skills for effective science communication: A teaching resource for undergraduate science education." International Journal of Science Education, Part B 7.2 (2017): 181-201.
AAAS Communication Toolkit: https://www.aaas.org/resources/communication-toolkit.
PLOS blog on scientific communication: https://scicomm.plos.org
Additional journal articles and online resources provided by the instructor.

RECOMMENDED PREREQUISITES
Basic communication in a professional setting.

KEYWORDS
Scientific communication.

Big Data Tools (MAG)
Outils des Big Data (MAG)

COURSE LANGUAGE

English in Marseille

TEACHER

Hervé MIGNOT – practitioner from Equancy

COURSE OUTLINE

1. Hadoop. HDFS. MapReduce. Stockage et calculs distribués. Déploiement d'un cluster.
2. Préparation, stockage et traitement des big data : Pandas, Hive and Pig
3. Data visualisation avec matplotlib & seaborn
4. Alternatives : solutions propriétaires, bases NoSQL, ElasticSearch

ORGANIZATION

Semester: S1
Teaching Hours: 24 h of lectures
Comment: Class exclusive for Magistere students.

Machine Learning and New Data (MAG)
Machine learning et nouvelles données (MAG)

COURSE LANGUAGE
English in Marseille

TEACHER
Quentin LIPPMANN – quentin.lippmann@univ-amu.fr

COURSE DESCRIPTION AND OBJECTIVES
This course proposes to study the processing and analysis of unstructured data, and more specifically textual data and images.

COURSE OUTLINE
This course is divided in two parts of 12 hours each. The first part covers text as data. The second is about image as data.

Part 1 – Text as Data

Block 1 – Foundations of NLP

Students will learn about the complete document-pre-processing pipeline, beginning with tokenisation and the construction of n-grams. They will create Bag-of-Words representations and apply TF–IDF weighting to highlight discriminative terms. We will then move to distributed word representations through word embeddings, extract entities with Named-Entity Recognition, and analyse sentence structure by performing dependency parsing.

Block 2 – Large Language Models

Students will learn about the transformer architecture and its self-attention mechanism, compare pre-training with fine-tuning, and experiment with in-context learning. They will study Reinforcement Learning from Human Feedback as an alignment method and practice prompt-engineering patterns to steer model behaviour. We will tackle hallucination and explore retrieval-augmented generation as a mitigation strategy.

Part 2 – Image as Data

Block 1 – Image Fundamentals

Students will learn about digital image representation and colour spaces, then examine convolution operations—kernel size, stride, padding—and their effect on the receptive field. They will study activation and pooling layers for feature extraction and understand bounding-box regression. Anchor-based and anchor-free object-detection strategies will be compared.

Block 2 – Facial Analysis, Segmentation & Generative AI

Students will learn about classical Haar cascades versus modern RetinaFace detectors for face localisation. They will map facial landmarks, build embedding-based recognition pipelines, and evaluate systems using FAR, FRR, ROC curves, and demographic-bias checks. Promptable segmentation models will be introduced, followed by diffusion-based generative models for image synthesis.

All the concepts are applied and illustrated in Python applications.

KEY PROFESSIONAL SKILLS UPON GRADUATION
To learn how to process and analyse textual data
To learn how to process and analyse images

ORGANIZATION
Semester: S1
Teaching Hours: 24 h of lectures
Comment: Class exclusive for Magistere students.

End-of-Studies Project (MAG)
Projet de fin d'études (MAG)

COURSE LANGUAGE

French in Marseille

TEACHER

A teacher + a practitioner

COURSE DESCRIPTION AND OBJECTIVES

The end-of-studies project is carried out in collaboration with a company from October to March. This project enables students to carry out operational engineering work in data science and to compare theory with applications in the professional world.

KEY PROFESSIONAL SKILLS UPON GRADUATION

To be able to tackle a data science problem and write a report to answer it.
To know how to work as a team and to meet a set of specifications.

ORGANIZATION

Semester: S1
Comment: Class exclusive for Magistere students. Bimonthly meetings with supervisors, and autonomous work between meetings.
Examination Method: Project + Defense

Topics in Data Science (MAG)
Sujets en Data Science (MAG)

COURSE LANGUAGE

English in Marseille

TEACHER

Pierre MICHEL – pierre.michel@univ-amu.fr
Christophe HURLIN – practitioner

COURSE DESCRIPTION AND OBJECTIVES

This course aims to raise students' awareness of topical issues in data science.

COURSE OUTLINE

1. Conformal prediction
   a. Introduction and theoretical foundations
   b. Conformal prediction for regression
   c. Conformal prediction for classification
2. Algorithmic fairness
   a. Introduction to algorithmic fairness
   b. Framework for fairness in machine learning
   c. Measuring algorithmic fairness
   d. Testing for algorithmic fairness
   e. Mitigating algorithmic biases

KEY PROFESSIONAL SKILLS UPON GRADUATION

Understand how to make conformal prediction for regression and classification
Understand algorithmic fairness, and how to measure it, test it and mitigate its biases.

ORGANIZATION

Semester: S2
Teaching Hours: 24 h of lectures
Comment: Class exclusive for Magistere students.

Projects in Data Science (MAG)
Projets en Data Science (MAG)

COURSE LANGUAGE

English in Marseille

TEACHER

Pierre MICHEL – pierre.michel@univ-amu.fr
Christophe HURLIN – practitioner

COURSE DESCRIPTION AND OBJECTIVES

This course is complementary to the course of « Topics in data science ». The goal of this course is to make students work on projects related to the topics studied in the other course.

KEY PROFESSIONAL SKILLS UPON GRADUATION

To be able to tackle a data science problem and write a report to answer it.

ORGANIZATION

Semester: S2
Teaching Hours: 24 h of lectures
Comment: Class exclusive for Magistere students.