COURSE OUTLINE

(1) GENERAL

SCHOOL	School of Finance and Statistics		
ACADEMIC UNIT	Department of Banking and Financial Management		
LEVEL OF STUDIES	Graduate		
COURSE CODE		SEMESTER	
COURSE TITLE	Statistical Learning in Finance		

INDEPENDENT TEACHING ACTIVITIES *if credits are awarded for separate components of the course, e.g. lectures, laboratory exercises, etc. If the credits are awarded for the whole of the course, give the weekly teaching hours and the total credits*	WEEKLY TEACHING HOURS	CREDITS
Add rows if necessary. The organisation of teaching and the teaching methods used are described in detail at (d).		

COURSE TYPE *general background, special background, specialised general knowledge, skills development*	General background
PREREQUISITE COURSES:	
LANGUAGE OF INSTRUCTION and EXAMINATIONS:	English
IS THE COURSE OFFERED TO ERASMUS STUDENTS	
COURSE WEBSITE (URL)	

(2) LEARNING OUTCOMES

Learning outcomes

The course learning outcomes, specific knowledge, skills and competences of an appropriate level, which the students will acquire with the successful completion of the course are described.

Consult Appendix A

- *Description of the level of learning outcomes for each qualifications cycle, according to the Qualifications Framework of the European Higher Education Area*
- *Descriptors for Levels 6, 7 & 8 of the European Qualifications Framework for Lifelong Learning and Appendix B*
- *Guidelines for writing Learning Outcomes*

This course aims to cover the basics of statistical learning so that students who successfully complete the course can have a smooth transition to the more advanced machine learning techniques. It starts with the core ideas regarding inference and prediction, supervised vs unsupervised learning, the bias-variance trade-ff, regression vs classification, and model assessment. It briefly reviews the linear regression and logistic regression models. Then, it introduces regularization and shrinkage ideas to develop the models of Ridge regression, Lasso and Elastic Net. It further covers dimension-reduction techniques such as Principal Components Analysis and Partial Least Squares. It further covers resampling techniques such as Cross-Validation and Bootstrap. Decision trees, such as Random Forests and Boosting, are also covered. Finally, we study Bayesian techniques, such as Naïve Bayes, Hierarchical Models, Markov Chain Monte Carlo, and Gaussian Processes.

Upon successful completion of the course, students should

- Be familiar with the main tools of Statistical Learning.
- Understand the importance of the bias-variance trade-off and the role of regularization.
- Understand the differences between supervised vs unsupervised learning and between regression and classification.
- Be able to decide which technique is more suitable for the empirical question at hand.
- Be able to implement the developed tools using Python and/or R.

General Competences

Taking into consideration the general competences that the degree-holder must acquire (as these appear in the Diploma Supplement and appear below), at which of the following does the course aim?

Search for, analysis and synthesis of data and information, with the use of the necessary technology	*Project planning and management*
Adapting to new situations	*Respect for difference and multiculturalism*
Decision-making	*Respect for the natural environment*
Working independently	*Showing social, professional and ethical responsibility and sensitivity to gender issues*
Team work	*Criticism and self-criticism*
Working in an international environment	*Production of free, creative and inductive thinking*
Working in an interdisciplinary environment	*……*
Production of new research ideas	*Others…*
	…….

Search for, analysis and synthesis of data and information, with the use of the necessary technology
Adapting to new situations
Decision-making
Working independently
Team work
Working in an interdisciplinary environment
Production of new research ideas

(3) SYLLABUS

1. Main Principles of Statistical Learning, Inference vs Prediction, Supervised vs Unsupervised Learning, Regression vs Classification, the Bias-Variance trade-off, in-sample vs out-of-sample performance, Regularization, Shrinkage, model assessment.
2. Bried review of the Linear Regression and the Logistic Regression models.
3. Introduction of the Bayesian Approach, Conjugate Priors, simple IID models, Linear Regression, Gibbs Sampling.
4. Regularization methods, Ridge Regression, Lasso, Elastic Net.
5. Dimension-reduction methods, Principal Components Analysis, Partial Least Squares.
6. Resampling methods, Cross-Validation and Bootstrap.
7. Naïve Bayes for Classification.
8. Gaussian Processes, Covariance Functions, Kernels, Hyperparameter Learning.
9. Decision Trees, Random Forest, Boosting.

(4) TEACHING and LEARNING METHODS - EVALUATION

DELIVERY *Face-to-face, Distance learning, etc.*	In-person
USE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY *Use of ICT in teaching, laboratory education, communication with students*	
TEACHING METHODS *The manner and methods of teaching are described in detail.* *Lectures, seminars, laboratory practice, fieldwork, study and analysis of bibliography, tutorials, placements, clinical practice, art workshop, interactive teaching, educational visits, project, essay writing, artistic creativity, etc.* *The student's study hours for each learning activity are given as well as the hours of non-directed study according to the principles of the ECTS*	see table below
STUDENT PERFORMANCE EVALUATION *Description of the evaluation procedure* *Language of evaluation, methods of evaluation, summative or conclusive, multiple choice questionnaires, short-answer questions, open-ended questions, problem solving, written work, essay/report, oral examination, public presentation, laboratory work, clinical examination of patient, art interpretation, other* *Specifically-defined evaluation criteria are given, and if and where they are accessible to students.*	Assessment will be based on (a) homework assignments and (b) a final project. Homework will be assigned every 3 weeks on average during the semester. For the homework assignments and the project students will be asked to apply the tools developed in class to real or simulated data using Python and/or R. Students can discuss the assignment questions with each other, but everyone must submit individual assignments.

Activity	*Semester workload*
Lectures	39
Independent Study	52
Empirical Analysis	59
Course total	**150**

(5) ATTACHED BIBLIOGRAPHY

- Suggested bibliography:

The Elements of Statistical Learning - Data Mining, Inference, and Prediction (2008) by Hastie, Tibshirani, and Friedman.

Gaussian Processes for Machine Learning (2008) by Williams and Rasmussen.

Pattern Recognition and Machine Learning (2011) by Bishop.

An Introduction to Statistical Learning with Applications in Python (2023) by James, Witten, Hastie, Tibshirani, and Friedman.

Machine Learning in Finance - From Theory to Practice (2020) by Dixon, Halperin and Bilokon.

A First Course in Statistical Learning (2025) by Lederer.