

Professional Training Course

Machine Learning: Tools and applications for policy



@EUI_FBF_School, @EUI_Schuman

6 - 8 June 2022

Sala Europa - Villa Schifanoia, Via Boccaccio 121 - Florence

Format: **Residential**

Course instructors:

- **Iman Van Lelyveld** | Vrije Universiteit Amsterdam and Dutch National Bank
- **Dieter Wang** | Vrije Universiteit Amsterdam and Dutch National Bank

EUI/BBF Team:

- **Thorsten Beck** | Scientific Advisor
- **Pierre Schlosser** | Scientific Advisor
- **María del Carmen Sandoval Velasco** | Research Associate
- **Chengzi Yi** | Teaching Associate

Introduction

It is hard to name a sector that will not be dramatically affected by machine learning or even artificial intelligence. There are many excellent courses that teach you the mechanics behind these innovations — helping you develop an engineering skill set. This course takes a different approach. It is aimed at people who want to deploy these tools, either in business or policy, whether through start-ups or within a large organization. While this requires some knowledge of how these tools work, it is only a small part of the equation — just as knowing how an engine works is a small part of understanding how to drive. What is really needed is an understanding of what these tools do well and what they do badly. This course focuses on giving you a functional, rather than mechanistic, understanding. By the end, you should be an expert at identifying ideal use-cases and thereby well-placed improve analysis and policy using machine learning.

This Florence School of Banking and Finance course thus aims to provide an overview and understanding of popular Machine Learning (ML) techniques, understand the opportunities and limitations of these techniques and be able to interact with the experts. In practical terms you will work hands-on with ML/AI methods in Python, demystify the black boxes and prime you so that you can continue to learn by yourself.



Programme

6 June

9.00 - 9.15 *Welcome coffee*

9.15 - 9.45 Presentation of the Florence School of Banking and Finance followed by a *tour de table*

9.45 - 10.30 **Lecture 1: Course Overview**

- Why is this course relevant?
- What can you expect?
- What will we cover?

10.30 - 11.00 *Coffee break*

11.00 - 12.00 **Tutorial 1: How to read data, how to use sklearn?**

- Getting started with Python and data manipulation.
- How is different from Excel?
- Read in the data and get to know it.
- Introduction to sklearn: where to find the buttons.

12.00 - 13.15 *Lunch break*

13.15 - 14.00 **Lecture 2: Introduction to Machine Learning (ML)**

- What is ML? What is ML applied to?
- Linear regression from the ML lens.
- The outlines of the ML approach:
 - Supervised unsupervised learning.
 - Hyperparameters and how to select them.
 - Gradient descent and grid search.

14.00 - 15.00 **Tutorial 2: Regressions versus Classifiers**

- Logit as a statistical model vs ML model.
- How to find the optimal (hyper)parameters.
- A different classifier: support vector machines
 - Different types of kernels.
 - First glimpse: dangers of overfitting.
 - Evaluating performance.

15.00 - 15.30 *Coffee break*

15.30 - 16.15 **Lecture 3: Machine Learning - the basics**

- Importance of preprocessing your data.
- Building up to the workhorse classifier: the logit model.
- When is a classifier doing a 'good' job?
- Confusion matrix, Receiver Operator Characteristic (ROC).
- What are overfitting, bias and variance?

16.15 - 17.15 **Tutorial 3: Data preprocessing and assessing model performance**

- How to preprocess: standardise your data.
- Pros and cons of standardisation.
- Working with the confusion matrix:
 - What if costs are not symmetric?
 - The trade-off between precision and recall.

17.15 - 17.45 Troubleshooting session with Teaching Associate (*optional*)

17.45 - 18.45 *Cocktail reception at the EUI*

7 June

9.00 - 10.15 **Lecture 4: Fighting the curse of dimensionality**

- How to reduce dimensionality?
 - K-Nearest Neighbors (KNN).
 - Principal Components Analysis (PCA).
- Feature selection and regularisation:
 - How to select the most important features?
 - Examples: RIDGE, LASSO, Elastic net.
- Is a 'good' model always good? What is external validity?
- Cross-validation and holdouts.

10.15 - 10.45 *Coffee break*

10.45 - 12.00 **Tutorial 4: Cross validation applied to LASSO variable selection**

- Looking closer at cross validation (CV) and holdouts.
- K-fold, Leave-one-out, stratified CV.
- Splitting your data into training and testing samples.
- How to use CV to tune a LASSO model.

12.00 - 13.15 *Lunch break*

13.15 - 14.00 **Lecture 5: Improving weak learning**

- How to grow a decision tree? How to split?
 - Purity measures.
- Can Ensemble Classifiers improve weak learners?
 - Bagging and boosting.
 - Examples: AdaBoost, XGBoost.

14.00 - 15.00 **Tutorial 5: Decision trees and random forests**

- Growing your own decision tree.
- How deep? How many splits? How big are the leaves?
- From trees to random forests.
- Comparing performance with the confusion matrix.

15.00 - 15.30 *Coffee break*

15.30 – 16.15 **Lecture 6: Unsupervised learning and clustering**

- Supervised versus unsupervised learning.
- What can we do with unsupervised learners?
- K-means, t-SNE, DBSCAN, Gaussian mixtures.

16.15 – 17.15 **Tutorial 6: Finding clusters and neighbors**

- Implementing K-means and DBSCAN.
- Hierarchical clustering: bottom-up or top-down?
- Visual inspection of results.

17.15 - 17.45 Troubleshooting session with Teaching Associate (*optional*)

19.30 Apericena at [B-Rooftop - Grand Hotel Baglioni](#) (Piazza dell'Unità Italiana, 6, 50123 Firenze)

8 June

9.00 – 9.45 **Lecture 7: Natural Language Processing (NLP)**

- What are the main approaches in textual analysis?
- Going beyond simple word counts.
- How to extract market sentiment?

9.45 - 10.45 **Tutorial 7: NLTK and sentiment analysis**

- Constructing a bag of words.
- Classifying sentiments (positive/negative).
- Example with financial news data.

10.45 - 11.15 *Coffee break*

11.15 - 12.00 **Lecture 8: Summary**

- Discuss some things that can go wrong:
 - Survivorship bias, input errors and deceit.
 - Fairness and discrimination.
- What is the reaction of authorities?

12.00 - 12.10 Concluding remarks

12.10 – 12.30 Troubleshooting session with the Teaching Associate (*optional*)