

# Machine learning for your workflow: Get the most from your data

## Online course

Centre for Econometric Analysis

Delivered by: Dr Jan Novotny

### Course overview

This course presents the machine learning field from the applied perspective. The objective is to give the audience an introduction into the field and show how to apply the presented methods in an accessible way. At the end of the workshop participants will be able to confidently apply presented methods to their own data set, making machine learning an inherent part of their daily work flow.

The course comprises of three modules, each of them building on the previous ones. The modules can be attended as a series, or, readers can choose of one or more modules according to their need and experience.

### Benefits

- You will be introduced to basic concepts of machine learning, which is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead
- You will learn how to include machine learning models into your daily workflow with practical examples
- You will learn how to code machine learning algorithms independently

- You will implement a number of different machine learning methods ranging from ordinary least squares, regularised linear regressions, decision trees and forest, boosting, to neural networks, understanding when each is applicable.

### Course prerequisites

You are expected to have a basic knowledge of Python (being able to run simple commands preferably using the Jupyter Notebooks), and a basic knowledge of Mathematics and Statistics.

### Target audience

This course is particularly useful to both professionals and researchers working in fields relevant, where there is demand for quantitative data-based decisions.

### Fees:

#### Single Module:

£120 Cass Alumni, Staff

£140 External Students

£240 External

#### Two Modules:

£180 Cass Alumni, Staff

£210 External Students

£360 External

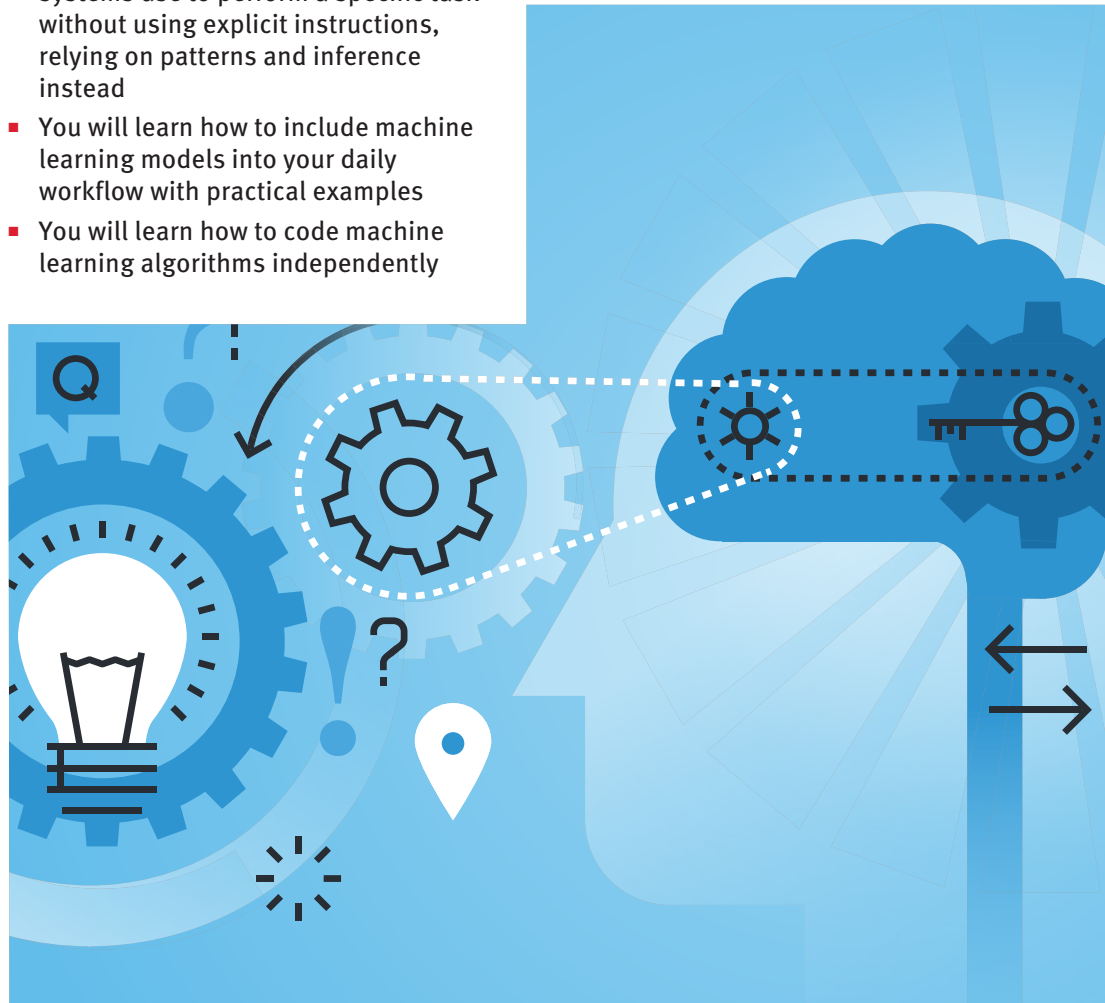
#### Three Modules:

£210 Cass Alumni, Staff

£245 External Students

£420 External

A 15% discount is available for groups of three or more participants





## Dr Jan Novotny

Jan Novotny (PhD, Charles University, Czech Republic) is an eFX Quant Trader at Deutsche Bank and research associate to the Centre for Econometric Analysis of Cass Business School in London. Prior his current role, he was a front office quant at HSBC in the electronic FX markets. Before joining HSBC team, he was working in the Centre for Econometric Analysis on the high-frequency time series econometric models and was visiting lecturer at Cass Business School, giving lectures at Warwick Business School and Politecnico di Milano. He has co-authored a number of papers in peer-reviewed journals in Finance (Journal of Financial Econometrics, Journal of Financial Markets) and Physics (Physica A, The European Physical Journal A), contributed to several books (Machine Learning and Big Data with kdb+/q, Wiley), 2019, and presented at numerous conferences and workshops world widely. During his PhD studies, he co-founded Quantum Finance CZ. He is a Machine Learning enthusiast and explores kdb+/q for this purpose.

### Module 1 – 4 hours online

#### Topic 1: Primer on probability and statistics and Introduction to data analysis

- Basics of the probability theory
- Review of elementary statistical concepts
- The basics of a data analysis
- Time series data and features like non-stationarity and memory.

#### Topic 2: Dimensionality reduction

- Dimensionality reduction
- Methods to locate significant feature in large datasets
- The Principal Component Analysis with applications
- Forward stepwise regression.

#### Topic 3: Linear regression

- The basic foundations of the ordinary least squares
- Accuracy of the fit
- Significance tests for the model itself and parameters
- Nested models.

#### Topic 4: Outlier detection

- The outlier detection in the data
- Basic principles of unsupervised machine learning method
- Implement the Local Outlier Factor
- Using PCA for outlier detection.

### Module 2 – 4 hours online

#### Topic 1: Bias-variance trade-off and regularised linear regression

- Balancing between the complexity of the model and its predictive accuracy
- The bias-variance trade-off
- Regularisation of machine learning problems and linear regressions
- The Ridge regression, and the Lasso regression.

#### Topic 2: Decision trees and random forests

- The binary decision tree and its practical considerations
- The CART method for both regression and classification
- The binary decision trees and its bootstrapped aggregates, or forests
- Complementary methods like boosting will be discussed.

### Topic 3: Working with large data sets

- The kdb+ database will be introduced and the basics of the q language
- Performing efficient operations on the vast datasets
- Illustrating the machine learning in q using quantQ library
- Comparing the tools for large data set manipulations.

### Topic 4: Unsupervised machine learning

- The unsupervised machine learning techniques
- Difference between supervised and unsupervised techniques
- Recap of the outlier detection as an unsupervised machine learning problem
- The Apriori algorithm building from scratch.

### Module 3 – 4 hours online

#### Topic 1: Boosting

- The boosting technique
- The AdaBoost
- Step-by-step improving of the algorithm design and stacking.

#### Topic 2: Neural networks and deep neural networks

- Introduction to the neural networks
- Regression and classification problems will be elaborated
- Discussing deep neural networks
- Link between neural networks, linear regressions and PCA.

#### Topic 3: Multi-arm bandit problem and Monte Carlo tree search

- The multi-arm bandit problem
- The Monte Carlo Tree Search
- Illustration of the MCTS concepts on the game of tic-tac-toe
- Discuss the famous Alpha Go (Zero) and possible use of MCTS in Finance.

## REGISTRATION, PAYMENT AND CANCELLATION POLICY

Payment of course fees is required prior to the course start date.

In case a course is cancelled, registered participants will receive the full refund.

Registration closes 7-calendar days prior to the start of the course.

## Recommended reading

The following textbooks and journal articles are recommended for this course:

Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second Edition February 2009 (<http://web.stanford.edu/~hastie/ElemStatLearn/>)

Kevin Sheppard, *Introduction to Python for Econometrics, Statistics and Numerical Analysis: Third Edition* ([https://www.kevinsheppard.com/Python\\_for\\_Econometrics](https://www.kevinsheppard.com/Python_for_Econometrics))

Jan Novotny, Paul Bilokon, Aris Galitos, Frederic Deleze, *Machine Learning and Big Data with KDB+/Q*, 2019, Wiley Finance Series

**Toolbox:** The readers are supposed to install the up-to-date version of Anaconda (<https://www.anaconda.com/>) and be able to run Jupyter Notebook with Python. In addition, the most recent version of q (free 32-bit version is fully sufficient) is recommended for Modules 2 and 3, where additional kdb+/q will be introduced.

