# Introduction to Machine Learning and Data Science for Finance

Online course

Centre for Econometric Analysis
Delivered by: Dr Jan Novotny

**Schedule:**

eight 90 minute long sessions, for 4 consecutive weeks (two sessions per week)

**Fees:**

**£360** City Students, Alumni, Staff

**£420** External Students

**£600** External

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

## Course overview

This course presents the machine learning field from an applied perspective. The objective is to give an introduction into the field and show how to apply the presented methods in an accessible way. At the end of the course participants will be able to confidently apply the methods to their own data set, making machine learning an inherent part of their daily work flow. The course comprises of **eight sessions,** each of them introducing a new topic and deepening the knowledge gathered in the previous ones. The course spans over **four weeks**, where every week, there are two sessions, each lasting 90 minutes. There will be a project assigned to delegates to practice the machine learning techniques, which will be discussed at the end. It is expected that delegates will further spend one/two hours a week to practice the machine learning techniques covered in the sessions.

## Benefits

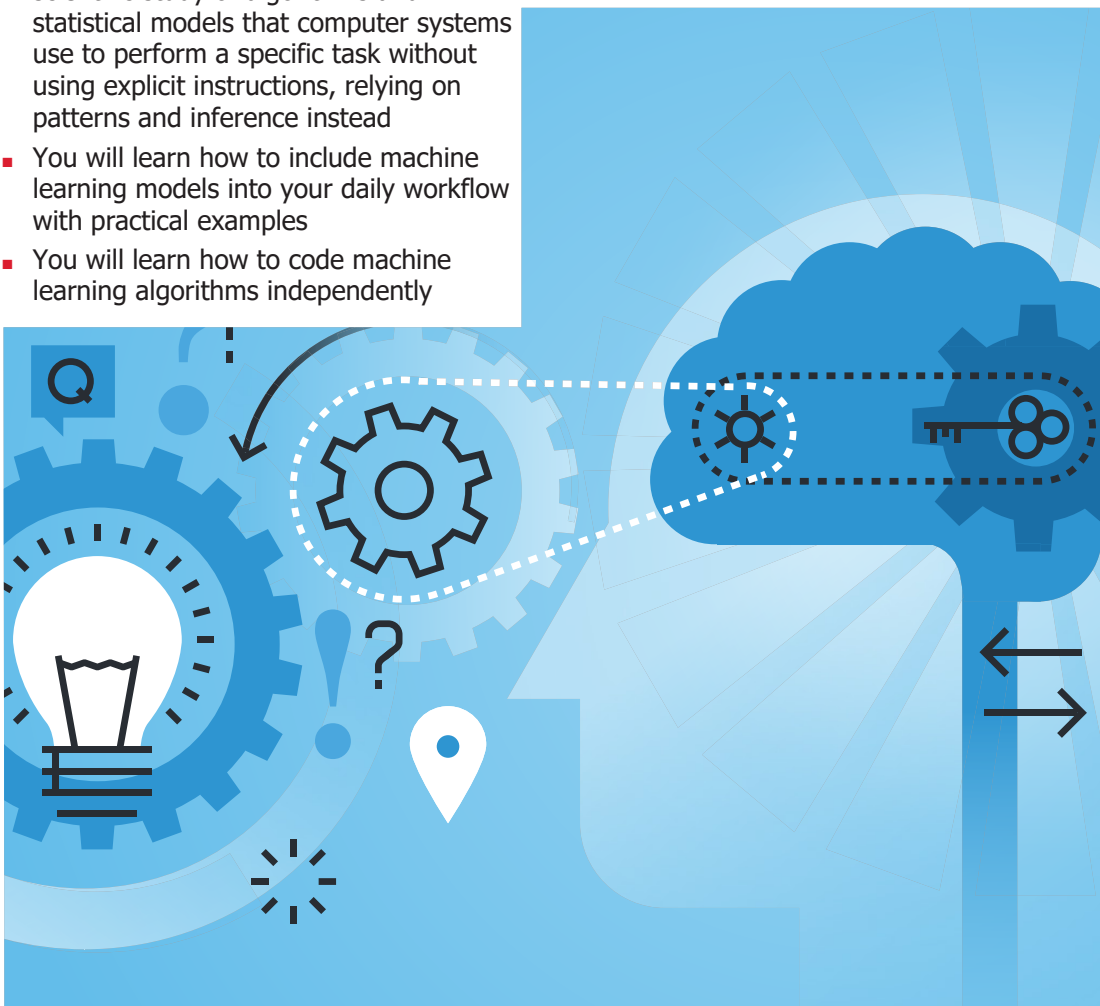
- You will be introduced to basic concepts of machine learning, which is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead
- You will learn how to include machine 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 neutral 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 where there is demand for quantitative data-based decisions.



For more info contact:
elisabetta. pellini@city.ac.uk
Cass Business School
CEA@city.ac.uk
www.cass.city.ac.uk/
cea-econometrics-courses

Dr Jan Novotny

Jan Novotny (PhD, Charles University, Czech Republic) is an eFX Quant at Nomura 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 Deutsche Bank and HSBC in the electronic FX markets. Before joining the industry, he was working at 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.

## Topic 1: Introduction to machine learning and data analysis

- Foundations of the Machine Learning
- Basics of the probability theory
- Review of elementary statistical concepts
- The basics of a data analysis
- Setting up computational tools

## Topic 2: Data, dimensionality reduction and project

- Dimensionality reduction
- The Principal Component Analysis with applications
- Forward stepwise regression
- Working with data, visualisations and data generation
- Project and the dataset

## Topic 3: Linear Regression and Regularised Linear Regression

- The basic foundations of the ordinary least squares
- Accuracy of the fit
- Significance tests for the model itself and parameters
- Nested models
- 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 4: Decision Tree, Random Forests and Model Stacking

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

## Topic 5: Unsupervised Machine Learning and Outlier Detection

- The unsupervised machine learning techniques
- Difference between supervised and unsupervised techniques
- Association rules using the Apriori algorithm and relevant metrics
- The outlier detection in the data: the supervised vs unsupervised Machine Learning approach
- The local outlier factor and the Principal Component Analysis

## Topic 6: Neural Networks

- The perceptron-based neural networks
- Classification and regression problems
- Auto-encoders and learning
- Discussing deep neural networks
- Link between neural networks, linear regressions and PCA

## Topic 7: Big data and kdb+/q

- The kdb+ database 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.
- The evaluation of the project

## Topic 8: Reinforcement Learning

- The multi-arm bandit problem
- The Monte Carlo Tree Search
- The illustration of the MCTS concepts on the game of tic tac-toe
- The famous Alpha Go (Zero) and use of MCTS in Finance
- Discussion: BYOP — Bring your own problem

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)

Jan Novotny, Paul Bilokon, Aris Galiotos, 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.