



Advanced AI for Data Analysis



Advanced AI for Data Analysis

The volume of data generated on the Internet is growing exponentially. Incredible amounts of unstructured data such as texts, images, time series, are available, all being produced at an overwhelming pace. In recent years, the availability of very high processing power (i.e., GPU's), combined with industrial interest, has given rise to new impressive AI methods, such as deep and reinforcement learning, which allow to extract significant information from large scale data in a very short time, solving highly complex problems and opening new perspectives. The Advanced AI Program is designed to provide deep insights on these new methods and their applications to massive and highly heterogeneous data. Important applications of these methods include marketing, gaming, recommendation systems, text mining, neuro-linguistic programming, social networks, fraud detection, image and video recognition, etc.



Target

Technical engineers, project managers, researcher and quantitative scientists.

Requirements: Skills and knowledge in data science (mathematics and machine learning) and programming (Python).



Objectives

This Program offers a comprehensive in-depth experience, with hands-on and presentation of the state-of-the-art AI techniques for large-scale data. The course covers deep learning for text, graph and time series mining; NLP; influence maximization; and recommendation algorithms.



Competencies

- › Acquire skills on recent machine learning methods
- › Master machine/deep learning tools and methodologies to address problems in text mining and NLP
- › Master machine/deep learning methods for graphs: link prediction, graph/node classification
- › Master AI algorithms to develop online marketing, fraud detection and knowledge extraction algorithms from web data
- › Integrate technical possibilities and ecosystem issues to characterize a data science-based project (massive data)
- › Analyze the opportunities, challenges and impacts associated with the use of data



Program



INTRODUCTION TO DATA MODALITIES AND LEGAL ASPECTS

Consumer analysis • Commercial offers • Client behaviour • Influencers • Legal aspects

DATA SCIENCE TOOLS - BIG DATA CONTEXT

Data base • Spark • Hadoop
Data science project pipeline • Exploration
Feature selection • Preprocessing
Dimensionality reduction

DEEP LEARNING

Introduction to advanced deep learning • Optimization of DL architectures • Attention based architectures • Transformers • Autoencoders for unsupervised learning

AI FOR TEXT MINING AND NLP

Architecture of web search engines • Advanced Machine learning for text and NLP • Deep learning methods for NLP • Word and Word/document and contextual word embeddings (ELMO, BERT) • Automated summarization • Entity recognition • Chatbots

AI FOR GRAPHS AND TIME SERIES

Node/graph embeddings • Graph kernels • Graph neural networks • Graph autoencoders • Deep sets

INFLUENCE MAXIMIZATION FOR SOCIAL/ COLLABORATION GRAPHS

SIR/SIS • Greedy algorithms • IMM (influence maximization via martingales) • Stop and stare kai SKIM
• Graph degeneracy based methods (D-core, RCG)

RECOMMENDATION ALGORITHMS

Factorization machines • NMF • Neural Collaborative Filtering • Deep Factorization Machines
Autoencoder based • Randomized SVD and SVD++.

DATA CHALLENGE

Potential topics: Opinion mining • Product recommendation • Link prediction • Chatbots...



Faculty director

MICHALIS VAZIRGIANNIS

École Polytechnique, Computer Science Department (DIX)



Certification École Polytechnique Executive Education

Lead a data science project using advanced AI tools



Application process

CV + cover letter



Duration

11 days + certification



Language

English



Location

École Polytechnique Executive Education, Palaiseau (available for corporate programs)



Contact

Mélanie Saint Rémy

melanie.saint-remy@polytechnique.edu

+33 (0)1 69 59 66 55



exed.polytechnique.edu

