Course : Big Data Analytics with Python

Practical course - 4d - 28h00 - Ref. BDA
Price : 2670 € E.T.

★★★★☆ 4,6 / 5

BEST

Teaching objectives

At the end of the training, the participant will be able to:

- ✓ Understanding the principle of statistical modeling
- ✓ Choosing regression and classification depending on data type
- ✓ Evaluating an algorithm's predictive performance
- ✓ Creating selections and classifications in large volumes of data to reveal trends

Practical details

Hands-on work

Developing/conducting analysis in Python, with the modules pandas, NumPy, SciPy, MatPlotLib, seaborn, scikit-learn, and statsmodels.

Course schedule

1 Introduction to modeling

- Introduction to the Python language.
- Introduction to the Jupiter Notebook software.
- Steps for building a model.
- Supervised and unsupervised algorithms.
- Choosing between regression and classification.

Hands-on work
Installing Python 3, Anaconda, and Jupiter Notebook.

2   Model evaluation procedures

- Techniques for resampling in training, validation and testing sets.
- Learning data representativeness test.
- Predictive model performance measurements.
- Confusion and cost matrix and AUC-ROC curve.

Hands-on work
Setting up data set sampling. Conducting evaluation tests on multiple provided models.

3   Supervised algorithms.

- The principle of univariate linear regression.
- Multivariate regression.
- Polynomial regression.
- Regularized regression.
- Naive Bayes.
- Logistic regression.

Hands-on work
Implementing regressions and classifications on multiple data types.

4   Unsupervised algorithms

- Hierarchical clustering.
- Non-hierarchical clustering.
- Mixed approaches.

Hands-on work
Handling unsupervised clusters in multiple datasets.

5   Component analysis

- Principal component analysis.
- Correspondence analysis.
- Multiple correspondence analysis.
- Factor analysis for mixed data.
- Hierarchical classification of principal components.

Hands-on work
Reducing the number of variables and identifying underlying factors of dimensions associated with significant variability.

6   Text data analysis

- Collecting and preprocessing text data.
- Extracting primary entities, named entities, and reference resolution.
- Grammatical tagging, syntactical analysis, semantic analysis.
- Lemmatization.
- Text vectorization.
- TF-IDF weighting.
- Word2Vec.

Hands-on work
Explore the contents of a text base using latent semantic analysis.

Dates and locations

REMOTE CLASS
2026 : 24 Feb., 21 Apr., 23 June, 23 June, 25 Aug., 25 Aug., 27 Oct., 27 Oct., 15 Dec.

PARIS LA DÉFENSE
2026 : 21 Apr., 23 June, 25 Aug., 27 Oct., 15 Dec.