Publication date : 01/19/2024

Course : NLP, natural language processing with Python

Natural Language Processing with Python tools and libraries

Practical course - 3d - 21h00 - Ref. PTS

Price : 1650 € E.T.

5 / 5

This course teaches the use of Python for natural language processing: data preparation, text representation and modeling. The participant uses Python tools and libraries to perform common NLP tasks, and implements and applies NLP models.

Teaching objectives

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

- Using python to process text data
- Select the Python tools and libraries needed for processing
- Set up the various preprocessing and vectorization stages
- Use appropriate techniques according to objectives: classification / topic modelling / sentiment analysis
- Apply and evaluate models on real data

Prerequisites

Programming skills in Python.

Course schedule

1 Python environment for NLP

- Python / Anaconda / Jupyter Notebook development environment.
- The main data types: strings, Booleans, numbers, lists, tuples and dictionaries.
- Control structures: for and while loops, if/elif/else tests.
- Functions: creation, parameter passing, default values, variable arguments.
- Numpy: vectors, matrices, slicing, concatenation.
- Pandas: analysis of tabular data (CSV, Excel...), statistics, pivots, joins, filters.

Hands-on work

Handling Python in a Jupyter notebook. Practical exercise with pandas and numpy.

PARTICIPANTS

PREREQUISITES

Programming skills in Python.

TRAINER QUALIFICATIONS

The experts leading the training are specialists in the covered subjects. They have been approved by our instructional teams for both their professional knowledge and their teaching ability, for each course they teach. They have at least five to ten years of experience in their field and hold (or have held) decision-making positions in companies.

ASSESSMENT TERMS

The trainer evaluates each participant's academic progress throughout the training using multiple choice, scenarios, hands-on work and more.
Participants also complete a placement test before and after the course to measure the skills they've developed.

TEACHING AIDS AND TECHNICAL RESOURCES

• The main teaching aids and instructional methods used in the training are audiovisual aids, documentation and course material, hands-on application exercises and corrected exercises for practical training courses, case studies and coverage of real cases for training seminars.
• At the end of each course or seminar, ORSYS provides participants with a course evaluation questionnaire that is analysed by our instructional teams.
• A check-in sheet for each half-day of attendance is provided at the end of the training, along with a course completion certificate if the trainee attended the entire session.

2 Text data pre-processing

- Identify what textual data are and introduce the spaCy and nltk libraries.
- Word tokenization.
- Removal of stop-words, punctuation and elements not essential to the analysis.
- Lemmatization vs. stemming.

Hands-on work

Preprocessing of text corpora with the 2 libraries, comparison of results and implementation methods. Creation of stop-word lists, comparison of lemmatization and rootization.

3 Information extraction

- Identify the grammatical nature of words using Part Of Speech Tagging.
- Identify people, places etc. with Named Entity Recognition.

Hands-on work

Implement Part Of Speech Tagging and Named Entity Recognition. Analysis of results, filters on certain grammatical categories and proper nouns.

4 Vector representation of text data

- Bag of words.
- Weighting tf-idf.
- Approach with n-grams.
- Embeddings: word2vec, gloVe, fastTesxt...

Hands-on work

Transformation of a text corpus using different approaches: bag of words, tf-idf, word2vec, gloVe. Comparison of vectors.

5 Machine learning on text data

- A reminder of the steps involved in building a predictive model.
- Classification.
- Sentiment analysis.
- Topic modelling.

Hands-on work

Modeling using different types of vectors (bag of words vs embeddings). Sentiment analysis on tweets.

6 Model evaluation procedures

- Resampling techniques in training, validation and test games.
- Testing the representativeness of training data.
- Performance measurements for predictive models.
- Confusion matrix.

Hands-on work

Build and evaluate an NLP model in an applied way...

TERMS AND DEADLINES

Registration must be completed 24 hours before the start of the training.

ACCESSIBILITY FOR PEOPLE WITH DISABILITIES

Do you need special accessibility accommodations? Contact Mrs. Fosse, Disability Manager, at psh-accueil@orsys.fr to review your request and its feasibility.

Dates and locations

REMOTE CLASS
2026 : 22 June, 2 Nov.

PARIS LA DÉFENSE
2026 : 22 June, 2 Nov.