Publication date : 10/02/2025

Course : Text mining in practice

Practical course - 3d - 21h00 - Ref. MMD
Price : 2010 € E.T.

Data mining restricted to textual data - text mining - is increasingly used in business. It can be used, for example, to rank products based on consumer feedback. You will apply text mining algorithms and tools to paradigmatic examples.

Teaching objectives

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

- Understanding textual statistics methods
- Implement feature extraction from textual data
- Create selections and rankings from large volumes of textual data
- Choosing a classification algorithm
- Evaluating the predictive performance of an algorithm

Intended audience

AI engineers/project managers, AI consultants and anyone interested in text mining for machine learning and deep learning.

Prerequisites

Good knowledge of statistics. Good knowledge of machine learning and deep learning. Experience required.

Course schedule

1 Traditional text mining approaches

- APIs for retrieving textual data.
- Preparing textual data according to the problem.
- Retrieval and exploration of the text corpus.
- Deleting accented and special characters.
- Stemming, lemmatization and removal of linking words.
- Gather everything together to clean up and standardize data.

Hands-on work

Document search, preparation, transformation and vectorization in DataFrame.

PARTICIPANTS

AI engineers/project managers, AI consultants and anyone interested in text mining for machine learning and deep learning.

PREREQUISITES

Good knowledge of statistics. Good knowledge of machine learning and deep learning. Experience required.

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.

2 Feature engineering for text display

- Understand text syntax and structure.
- The Bag of Words and Bag of N-Grams models.
- The TF-IDF model, Transformer and Vectorizer.
- The Word2Vec model and implementation with Gensim.
- The GloVe model.
- The FastText model.

Hands-on work
Set up operations to extract features from textual data in order to carry out classifications.

3 Text similarity and unsupervised classification

- The essential concepts of similarity.
- Term similarity analysis: Hamming, Manhattan, Euclidean and Levenshtein distances.
- Document similarity analysis.
- Okapi BM25 and the ranking list.
- Unsupervised classification algorithms.

Hands-on work
Build a recommendation system for similar products based on the description and content of the products you've chosen.

4 Supervised text classification

- Data pre-processing and normalization.
- Classification models.
- Multinomial Naive Bayes.
- Logistic regression. Support Vector Machines.
- Random Forest. Gradient Boosting Machines.
- Evaluation of classification models.

Hands-on work
Implementation of supervised classifications on multiple datasets.

5 Natural Language Processing and deep learning

- NLP libraries: NLTK, TextBlob, SpaCy, Gensim, Pattern, Stanford CoreNLP.
- Deep learning libraries: Theano, TensorFlow, Keras.
- Natural Language Processing and Recurrent Neural Networks.
- RNN and Long Short-Term Memory. Bidirectional RNN models.
- Sequence-to-Sequence models.
- Questions and answers with RNN models.

Hands-on work
Build an RNN to generate new text.

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.

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 **PARIS LA DÉFENSE**

2026 : 1 June, 12 Oct.

2026 : 1 June, 12 Oct.