

Training - Text Analysis using Machine Learning



MLTXT-EN-D



Distance only



1 day

In the Oil & Gas industry, like in other domains, most of the engineering knowledge and experience is saved in unstructured documents such as PDF files, MS Office files or just scans of paper reports. Extracting information from this file required a tremendous effort from domain experts and data officers since deterministic data mining tools have limited capabilities. Today some new machine learning systems propose some new ways to extract more information from unstructured documents. This course will review the more common used algorithms in the domain of text analysis and is a good introduction to the domain of machine learning.

Level

Awareness

Public

Geoscientists who have to use machine learning and/or text analysis tools to reach their business objectives (Boot camp training with a notebook designed for people having no or limited python experience, but with general knowledge of coding)

Objectives

Attendees will be able to implement the following skills:

- Understand the main algorithms used for text mining and text analysis, and run a text classifier using a Colab notebook

Pedagogical & technical resources

This course can be delivered as a presential or virtual classroom. Each training module contains lectures, hands-on practices and/or case studies

Assessment of achievements

- Trainees are assessed throughout the training through practical application phases and interactions with the trainer
- A final on-the-spot evaluation may also be carried out at the end of the course and/or at the end of each module using tests designed to verify the learners' understanding and assimilation of the knowledge linked to the training objectives

Prerequisites

No prerequisites are necessary to follow this course

Responsible

IFP Training instructors, with expertise in the field and trained in modern teaching methods adapted to the specific needs of learners from the professional world

Program

INTRODUCTION

Problem statement.

TEXT CLASSIFICATION USING A ML ALGORITHM IN A NOTEBOOK

Exploring the document with EDA.

Features engineering using the scikit-learn library.

Model training:

- Logistic regression.

- Passive aggressive classifier.
- Random forest classifier.
- Gradient boosting classifier.

Model validation:

- Precision, recall, F1 score.
- Confusion matrix.
- Lime, ELI5.

NATURAL SPEECH PROCESSING

Semantic analysis.

Markov chains.

Sentiment analysis.

DETECTING KEY-WORDS/METADATA VALUES IN TEXT

Detecting candidate values with deterministic method.

ML to associate a confidence factor to each candidate.

Improving the model thanks to the user experience.

To French entities : IFP Training is referenced to DataDock ; you may contact your OPCO about potential funding.

Please contact our disabled persons referent to check the accessibility of this training program : referent.handicap@ifptraining.com