

NLP with Python

Nom : **PICAULT**

Prénom : **Matthieu**

Année : **M2**

Semestre : **9**

Nature : **CM**

Volume horaire : **12 H**

ECTS / Coef : **2**

Prérequis	Introduction à Python (M1 ESA) Installation de la suite Anaconda
Résumé	<p>Ce cours se concentre sur le traitement des informations textuelles (articles de presse, réseaux sociaux, documents officiels, ...) et son utilisation dans les domaines de l'Economie et de la Finance.</p> <p>Afin d'apprendre à extraire les informations importantes d'un texte, les étudiants s'appuient sur les principaux packages de Python en analyse textuelle comme <i>Nltk</i> et <i>Spacy</i> et travaillent à partir de cas pratiques.</p> <p>L'évaluation s'appuie sur un projet de groupe.</p>
Objectifs	<p>L'objectif de ce cours est de permettre aux étudiants de :</p> <ul style="list-style-type: none">- Manipuler des données textuelles (<i>string</i>).- Isoler les éléments important d'un texte en étudiant sa structure grammaticale (<i>Part-of-Speech tagging</i>)- Calculer la complexité de documents au sein d'un corpus et mesurer la similarité entre les documents.- Identifier les occurrences (<i>TF-IDF</i>) et les thèmes principaux d'un corpus de documents (<i>Latent Dirichlet Allocation</i>).- Estimer le sentiment associé aux documents du corpus.- Introduction à l'utilisation des Machine Learning supervisés aux données textuelles.
Bibliographie	<ul style="list-style-type: none">• Bird, S., Klein, E., & Loper, E. (2009). <i>Natural language processing with Python: analyzing text with the natural language toolkit</i>. " O'Reilly Media, Inc."• Hardeniya, N., Perkins, J., Chopra, D., Joshi, N., & Mathur, I. (2016). <i>Natural language processing: python and NLTK</i>. Packt Publishing Ltd.• Perkins, J. (2014). <i>Python 3 text processing with NLTK 3 cookbook</i>. Packt Publishing Ltd.• Beysolow II, T. (2018). <i>Applied Natural Language Processing with Python: Implementing Machine Learning and Deep Learning Algorithms for Natural Language Processing</i>. Apress.• Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. <i>Journal of Accounting Research</i>, 54(4), 1187-1230.

PLAN

Introduction: What is NLP and text analysis?

Section 1: Preparing a corpus and the NLP pipeline

Section 2: Similarity, complexity, and sentiment.

Section 3: Machine Learning using text as data

Section 4: Identifying the topics across a corpus using the LDA method.