



# NLP : Natural Language Processing

Ce cours décrit les outils et techniques nécessaires pour l'exploration, le traitement et l'analyse des données textuelles. Il montre comment appliquer des techniques d'apprentissage automatique (machine learning) et d'analyse de données sur du texte pour le transformer en données compréhensibles par machine, et appliquer des opérations telles que l'analyse des sentiments, la détection de spam, la traduction automatique, etc. Ce cours se concentre sur l'application de techniques d'apprentissage profond (Deep Learning), sans aller profondément dans la théorie mathématique mais plutôt en utilisant des modèles pré-implémentés pour les tâches du NLP (Natural Language Processing)

## Programme

### FOCUS 1 Introduction au NLP

C'est quoi le NLP, Applications du NLP, Défis du NLP

### FOCUS 2 Prétraitement de texte

Collecte de données (web scraping, social media, open data sources)  
Tokenization, stopwords, stemming, lemmatization

### FOCUS 3 Extraction d'information

Part-of-speech (POS) tagging  
Named-entity recognition (NER)

### FOCUS 4 Term frequency inverse document frequency (TFIDF)

bag-of-words, vectorization, term frequencies, vector space in NLP, document similarity, calculation, Zipf's law, inverse document frequency

### FOCUS 5 Word Embeddings

word vectors, vector-oriented reasoning, Word2Vec  
other word embedding libraries (GloVe, FastText), visualization, document similarities with Word2Vec

### FOCUS 6 Neural Networks

Introduction, the perceptron, back propagation

### FOCUS 7 Convolutional Neural Networks (CNN)

Introduction, filters, padding, convolution in NLP

### FOCUS 8 Recurrent Neural Networks (RNN)

CNN vs RNN, recurrence in neural networks, applications of RNN

### FOCUS 9 Long Short-Term Memory (LSTM)

Introduction, LSTM structure, LSTM applications

## Objectifs

- Comprendre les différentes applications de l'analyse de texte et le rôle du NLP
- Apprendre à collecter et à traiter des données textuelles brutes et à les transformer en informations interprétables par la machine
- Apprendre le concept du word embedding et les différentes transformations du texte dans l'espace vectoriel
- Appliquer la régression logistique et la méthode Bayésienne pour analyser des sentiments, résumer de texte et traduire des mots
- Utilisez des réseaux de neurones récurrents (RNN) et des LSTM pour l'analyse des sentiments, la génération de texte et la reconnaissance d'entités nommées

# NLP : Natural Language Processing

## Pédagogie et ressources

La formation s'articule autour d'exposés, d'ateliers pratiques et de partages de retours d'expérience.

## Modalités d'évaluation

Évaluation des acquis tout au long de la session par des travaux pratiques.

## Moyens techniques

En fonction du format, distanciel via l'outil Teams, en présentiel, salle de formation équipée de postes de travail informatiques disposant de tous les logiciels nécessaires au déroulement de la formation, salle moderne climatisée, accès à l'environnement numérique Efrei.

## Prérequis

Connaissance basique du Python.

Connaissance basique du machine learning et réseaux de neurones.

## Profil du participant

Ingénieurs ou chef de projet data/IA ayant une première expérience en machine learning et en Python.

## Sanction de la formation

Une attestation de fin de formation résumant les objectifs visés par la formation est remise au participant à l'issue de la formation.

## Prix

2 100€ HT\* par participant

## Direction pédagogique

Salim NAHLE

## Contact

[executive.education@efrei.fr](mailto:executive.education@efrei.fr)

06.23.18.43.22

\* Prix HT, les déjeuners des jours de formation sont inclus.  
Prix, dates, équipes pédagogiques et contenu des programmes sont susceptibles de changer.  
Délai d'accès entre 3 et 5 jours ouvrés  
Accessibilité : <https://www.efrei.fr/ecole-ingénieur/efrei-for-good/>  
Contact : handicap@efrei.f



## Format

Présentiel ou distanciel (via l'outil Teams)



## Durée

3 Jours (21 heures)



## Dates

1<sup>ère</sup> session : 12 au 14 Juin 2024

2<sup>ème</sup> session : 7 au 9 Octobre 2024

**INSCRIPTION**