



PROGRAMME DE FORMATION SUR TRAITEMENT AUTOMATIQUE DU LANGAGE NATUREL (NLP)

Description de la formation

Cette formation en Traitement Automatique du Langage Naturel (TALN) vous initie aux méthodes modernes d'analyse et de modélisation des textes avec Python. Elle couvre les bases du prétraitement, la vectorisation (TF-IDF, Word2Vec), l'analyse de sentiment, la classification de textes et l'extraction d'informations. Des modules avancés portent sur les réseaux de neurones séquentiels (LSTM, GRU) et les modèles Transformers (BERT, GPT). Vous apprendrez à utiliser les bibliothèques Python telles que NLTK, spaCy, Gensim et Hugging Face Transformers. Des cas pratiques incluent la création de chatbots, la génération de texte et le résumé automatique. La formation est basée sur des travaux dirigés, des exercices pratiques et un projet final. Elle s'adresse aux étudiants, professionnels et chercheurs ayant des bases en Python et machine learning. Un accent est mis sur le déploiement des modèles NLP via des API simples. À la fin, vous serez capable d'appliquer le NLP à des problèmes réels. Une attestation ou un certificat est délivré après évaluation finale.



MODULE 1 : Introduction au NLP et au traitement du texte brut

- Définition, historique et applications du NLP
- Installation des outils : Python, Jupyter, NLTK, spaCy
- Traitement de texte brut : nettoyage, tokenisation, stopwords
- Normalisation : stemming vs lemmatisation
- **TP** : Prétraitement de documents textuels (corpus d'articles)

MODULE 2 : Représentation vectorielle du texte

- Modèles de sac de mots (BoW), TF-IDF
- N-grams, matrice de co-occurrence
- Introduction aux embeddings : Word2Vec, GloVe
- Visualisation avec PCA, t-SNE
- **TP** : Classification de texte avec BoW et TF-IDF

MODULE 3 : Analyse textuelle classique

- Classification de texte : SVM, Logistic Regression, Naive Bayes
- Analyse de sentiment
- Extraction d'entités nommées (NER)
- Détection de topics : LDA (Latent Dirichlet Allocation)
- **TP** : Classifier des avis clients ou des tweets



MODULE 4 : Modèles séquentiels avec RNN, LSTM, GRU

- Introduction aux réseaux de neurones séquentiels
- Entraînement de modèles RNN et LSTM avec Keras ou PyTorch
- Encodage des séquences, padding, masking
- **TP** : Génération automatique de texte ou analyse de sentiments avec LSTM

MODULE 5 : NLP moderne avec Transformers

- Introduction aux Transformers, BERT, GPT, etc.
- Hugging Face Transformers : chargement et fine-tuning de modèles pré-entraînés
- Applications : Q&A, résumé, classification, traduction
- **TP** : Fine-tuning de BERT pour la classification d'e-mails ou d'avis

MODULE 6 : Chatbots et applications NLP avancées

- Architecture d'un chatbot NLP
- Rasa, Dialogflow ou simple chatbot avec Python
- Résumé automatique, génération de texte, traduction automatique
- **TP** : Création d'un chatbot simple avec règles et NLP



MODULE 7 : Déploiement et industrialisation

- Sauvegarde des modèles NLP
- Déploiement avec Flask, Streamlit ou FastAPI
- Evaluation des modèles : précision, F1, BLEU score, perplexité
- **TP** : Création d'une API de classification de texte

MODULE 8 : Projet final

- Choix du projet : analyse de sentiment, résumé automatique, chatbot, etc.
- Développement, test, documentation
- Soutenance devant jury ou instructeur
- **Livrable** : Code, rapport technique, présentation

Outils et librairies utilisées :

- **Langage** : Python
- **Librairies** : NLTK, spaCy, Scikit-learn, Gensim, Keras, PyTorch, Transformers (Hugging Face), Flask
- **Environnements** : Jupyter, Google Colab, VS Code

Prérequis :

- Connaissances de base en Python
- Connaissances en machine learning et statistiques
- Notions de traitement de données (Pandas, NumPy)