

Plan d'Apprentissage : Fine-Tuning Transformers pour l'Aspect-Based Sentiment Analysis (ABSA) en E-Tourisme

Plan d'Apprentissage (4-5h par jour)

1 Collecte et Prétraitement des Données

Objectifs : Apprendre à collecter et prétraiter des données textuelles pour la tâche ABSA.

Théorie :

- Introduction au *web scraping* et aux APIs (Twitter, TripAdvisor, Booking.com).
- *Prétraitement des données textuelles* : Tokenization, Stopwords, Stemming, Lemmatization.
- Gestion des *imbalances de classes* dans les datasets.

Pratique :

- Récupérer un dataset d'avis clients depuis Kaggle ou via scraping.
- Appliquer un pipeline de prétraitement avec `spaCy` ou `NLTK`.
- Nettoyer les textes et gérer les données manquantes.

Certifications recommandées pour cette étape

- **Using Python to Access Web Data** – Coursera (University of Michigan)
Lien : <https://www.coursera.org/learn/python-network-data>
- **Feature Engineering for NLP in Python** – Datacamp
Lien : <https://www.datacamp.com/courses/feature-engineering-for-nlp-in-python>

2 Introduction au NLP et à l'ABSA

Objectifs : Acquérir les bases du NLP et de l'Aspect-Based Sentiment Analysis (ABSA).

Théorie :

- Introduction au *NLP* (Natural Language Processing).
- Exploration des tâches fondamentales : Tokenization, POS tagging, NER, etc.
- Définition de l'ABSA et comparaison avec l'analyse de sentiment classique.

Pratique :

- Lecture d'articles sur l'ABSA en e-tourisme et analyse d'exemples d'avis classés (*positive*, *negative*, *neutral*).

Certifications recommandées pour cette étape

- **Natural Language Processing Specialization** – Coursera (DeepLearning.AI)
Lien : <https://www.coursera.org/specializations/natural-language-processing>

3 Modèles Classiques en NLP et Introduction aux Transformers

Objectifs : Comprendre les modèles classiques avant l'avènement des Transformers.

Théorie :

- *Modèles traditionnels* en NLP : TF-IDF, Bag of Words, Word2Vec.
- Introduction aux *Transformers* et leurs concepts fondamentaux (Self-attention, Positional Encoding).
- Présentation des modèles BERT, RoBERTa, T5.

Pratique :

- Comparaison de TF-IDF et Word2Vec sur un dataset d'avis clients.
- Visualisation des représentations vectorielles avec Word2Vec.

Certifications recommandées pour cette étape

- **Natural Language Processing Fundamentals in Python** – DataCamp
Lien : <https://www.datacamp.com/courses/natural-language-processing-fundamentals-in-python>

4 Introduction aux Transformers et Hugging Face

Objectifs : Découvrir l'architecture des Transformers et apprendre à utiliser la bibliothèque Hugging Face.

Théorie :

- Fonctionnement des *Transformers* : Self-attention et Multi-head Attention.
- Présentation de *Hugging Face* et de sa bibliothèque pour exploiter des modèles pré-entraînés.

Pratique :

- Tester des modèles pré-entraînés avec Hugging Face (ex : BERT pour la classification des sentiments).
- Utilisation d'un tokenizer pour encoder un dataset.

Certifications recommandées pour cette étape

- **Transformers for NLP** – Datacamp
Lien : <https://www.datacamp.com/courses/transformers-for-nlp>

5 Fine-Tuning d'un Modèle Transformer pour l'ABSA

Objectifs : Apprendre à fine-tuner un modèle Transformer pour la tâche ABSA.

Théorie :

- Différence entre *fine-tuning* et feature extraction.
- Adaptation d'un modèle pré-entraîné pour une tâche ABSA spécifique.
- Préparation des données d'entraînement.

Pratique :

- Fine-tuning du modèle BERT sur un dataset ABSA.
- Ajustement des hyperparamètres pour optimiser les performances.

Certifications recommandées pour cette étape

- **Fine-tuning BERT with Hugging Face** – Coursera
Lien : <https://www.coursera.org/learn/fine-tuning-bert-hugging-face>

6 Cas Pratique Simple (ABSA)

Objectifs : Mettre en pratique le fine-tuning pour une tâche ABSA concrète.

Pratique :

- Sélection d'un dataset d'avis clients (ex : SemEval 2014).
- Application du fine-tuning sur le dataset et évaluation des performances (accuracy, precision, recall).
- Analyse des erreurs et optimisation des hyperparamètres.

Certifications recommandées pour cette étape

- **Applied Text Mining in Python** – Coursera (University of Michigan)
Lien : <https://www.coursera.org/learn/python-text-mining>

7 Finalisation et Optimisation du Modèle

Objectifs : Optimiser le modèle, analyser les résultats, et finaliser le projet.

Pratique :

- Optimisation du modèle pour atteindre des performances maximales.
- Analyse des résultats sur un jeu de données de test.
- Ajustement des hyperparamètres et gestion des erreurs.
- Déploiement du modèle via une API ou une interface utilisateur.
- Rédaction du rapport final du projet.

Certifications recommandées pour cette étape

- **Machine Learning Engineering for Production (MLOps)** – Coursera (DeepLearning.AI)
Lien : <https://www.coursera.org/specializations/machine-learning-engineering-for-production-mlops>