**Étape suivante après le "Data Preprocessing" :**

1. **Exploration des algorithmes de ML :** Avant de plonger dans le deep learning, il est essentiel de maîtriser certains algorithmes classiques de machine learning. Ces algorithmes sont souvent utilisés comme bases de comparaison pour évaluer des modèles plus complexes.
2. **Maîtrise des techniques d'évaluation :**
   * Validation croisée
   * Courbes ROC et métriques comme précision, rappel, F1-score
   * Matrice de confusion (très utile pour classifier correctement les données NLP)

**Algorithmes de ML obligatoires à réviser :**

Ces algorithmes sont fondamentaux et apparaissent dans presque tous les domaines, y compris en NLP.

1. **Régression :**
   * Régression linéaire (pour comprendre la base des prédictions continues)
   * Régression logistique (utile pour des tâches de classification)
2. **Arbres de décision et Forêts aléatoires :**
   * **Arbres de décision** : pour comprendre comment les données peuvent être partitionnées.
   * **Random Forest** : un modèle puissant basé sur l'ensemble des arbres.
3. **Support Vector Machines (SVM) :**
   * Très utile pour les tâches de classification, y compris en NLP comme la classification de texte.
4. **K-Nearest Neighbors (KNN) :**
   * Simple mais efficace pour certains types de données.
5. **Clustering :**
   * **K-Means** : utile pour des tâches d'exploration non supervisée.
   * **DBSCAN** : pour travailler avec des données bruitées.
6. **Naïve Bayes :**
   * Très important en NLP pour des tâches comme la classification de texte et l'analyse de sentiments.
7. **Gradient Boosting :**
   * **XGBoost** ou **LightGBM** : souvent utilisé pour des compétitions ML et pour des performances élevées sur des données tabulaires.
8. **Réseaux de Markov cachés (HMM) :**
   * Très pertinents en NLP pour des tâches comme l'analyse de séquences (POS tagging, etc.).

**Révision de base avant le Deep Learning :**

Il est essentiel d’avoir une bonne compréhension des concepts suivants pour aborder le deep learning efficacement :

1. **Principes mathématiques :**
   * Algèbre linéaire (vecteurs, matrices, multiplications matricielles)
   * Calcul différentiel (pour comprendre la descente de gradient)
   * Probabilités et statistiques
2. **Concepts ML avancés :**
   * **Overfitting et Underfitting** : comprendre comment gérer ces problèmes.
   * **Regularisation** : L1, L2, et dropout (pour le deep learning).
   * **Ensemble Learning** : bagging et boosting.

**Étape suivante après ML : Introduction au Deep Learning**

Une fois à l’aise avec ML, commence doucement le deep learning :

1. **Introduction au DL :**
   * Qu’est-ce qu’un réseau de neurones ?
   * Fonctionnement de la propagation avant (forward propagation) et arrière (backpropagation).
   * Compréhension des fonctions d'activation comme ReLU, Sigmoid, Tanh, etc.
2. **Bibliothèques populaires :**
   * **TensorFlow** ou **PyTorch** : commence par une introduction pour manipuler les tenseurs.
3. **Types de réseaux :**
   * **Réseaux de neurones denses (ANN)** : pour comprendre les bases.
   * **Réseaux convolutifs (CNN)** : pour des tâches liées aux images (utile pour comprendre les bases avant les applications NLP avancées comme Transformers).
   * **Réseaux récurrents (RNN)** et **LSTM/GRU** : nécessaires pour NLP.

**Plan d'apprentissage NLP :**

Après avoir révisé ML et commencé DL, voici ce que tu peux aborder pour NLP :

1. **Tokenization et Vectorization :**
   * Comprendre comment transformer du texte brut en vecteurs (Bag of Words, TF-IDF, Word Embeddings).
   * Utiliser des embeddings comme Word2Vec, GloVe, ou FastText.
2. **Modèles classiques pour NLP :**
   * Naïve Bayes, SVM pour la classification de texte.
   * HMM et CRF pour l'analyse de séquences.
3. **Transformer-based models :**
   * Introduction aux modèles basés sur Transformers (comme BERT et GPT).
   * Hugging Face est un excellent point de départ.