

# Rapport détaillé - Reconnaissance d'entités nommées avec LSTM et variantes

---

## Introduction

Ce projet a pour objectif de réaliser une tâche de reconnaissance d'entités nommées (Named Entity Recognition - NER) à l'aide de modèles à base de réseaux de neurones récurrents (LSTM, GRU) et convolutionnels (CNN). Nous utilisons le corpus CONLL-2003 et les vecteurs de mots pré-entraînés GloVe.

## Approche

- Préparation des données : Lecture, nettoyage, analyse exploratoire.
- Représentation vectorielle : Embeddings GloVe 100d figés.
- Encodage : Mapping mots/index + padding + one-hot encoding des étiquettes.

## Architectures Testées

- BiLSTM + Dropout (modèle de base)
- CNN + BiLSTM
- GRU + Dropout
- LSTM simple + Dropout
- LSTM + couche Dense intermédiaire

## Analyse

- Très bonne performance sur les classes majoritaires (O).
- Les classes minoritaires sont plus complexes à apprendre.
- L'utilisation des embeddings GloVe a fortement aidé la performance globale.

## Conclusion

Nous avons conçu et comparé plusieurs architectures pour la tâche de NER. Le meilleur compromis précision/complexité est obtenu avec le modèle BiLSTM simple avec Dropout, atteignant une macro F1 de 84%.

## Résultats et Comparaison

Modèle	Accuracy	Macro F1-score
BiLSTM + Dropout	99%	84%
CNN + BiLSTM	99%	83%
GRU + Dropout	99%	83%
LSTM simple + Dropout	99%	78%
LSTM + Dense	99%	79%