



Stage de Master : Deep Learning pour l'analyse de sentiment dans des témoignages écrits.

Localisation: Laboratoire LITIS, Université de Rouen Normandie, Rouen, France (www.litislab.fr)

Durée: 6 mois avec gratification

Période : entre février et décembre 2022

Mots-clés : *Deep Learning*, Traitement Automatique de la Langue (NLP), analyse de sentiment, *Transformers*

L'équipe Apprentissage du laboratoire LITIS, le laboratoire d'informatique et d'intelligence artificielle de l'Université de Rouen Normandie, propose un stage de 6 mois de fin de Master ou de fin de cursus Ingénieur, à débiter en février/mars 2022. Ce stage est financé par le projet ANR CATCH (Compréhension Automatique de Témoignage de Capteurs Humains) qui implique le centre Recherche et Développement de l'entreprise Saagie¹, une entreprise spécialisée dans les solutions DataOp B2B, Atmo Normandie², l'association de surveillance de la qualité de l'air en Normandie, et le LITIS.

Contexte Scientifique :

L'ambition du projet CATCH est de proposer des outils d'intelligence artificielle et de *deep learning* pour identifier et analyser automatiquement la multitude de témoignages humains liés à un accident industriel et à ses conséquences sur l'environnement et la santé. En impliquant la population dans la collecte et l'analyse des données, notamment via les réseaux sociaux, et en fournissant des moyens efficaces d'interprétation de ces données, la solution proposée devrait contribuer à apporter des réponses à la problématique préoccupante des accidents industriels et de leurs conséquences.

Pour cela, l'objectif du projet est double :

1. Dresser une cartographie des nuisances dues à l'incident, afin de suivre la propagation et l'évolution des phénomènes dans le temps.
2. Analyser et caractériser le ressenti de la population et son évolution tout au long de la crise

Pour ce faire, nous pouvons exploiter les témoignages recueillis sur la plateforme ODO³ d'Atmo Normandie, qui combine ces témoignages avec des informations géographiques, ainsi que des données extraites de la plateforme de micro-blogging Twitter.

Objectif du stage :

Ce stage vise à contribuer à ces deux objectifs via l'analyse des sentiments exprimés dans les témoignages de ces données, déjà disponibles et annotées en ce qui concerne les données ODO, ou à annoter automatiquement pour les données Twitter.

Les méthodes à l'état de l'art pour ce type de tâches en traitement automatique de la langue (TAL ou NLP pour *Natural Language Processing*) sont des méthodes d'apprentissage profond de type *Transformers*⁴, notamment basées sur des principes d'apprentissage auto-supervisés (*self-supervised*

¹ <https://www.saagie.com/>

² <http://www.atmonormandie.fr/>

³ <https://www.atmo-odo.fr/>

⁴ <https://super.gluebenchmark.com/leaderboard>

learning) et des mécanismes d'attention⁵. Donc, la première tâche de ce stage consistera à sélectionner, implémenter et appliquer une ou plusieurs de ces approches pour l'analyse de sentiment dans les données ODO. Il s'agira ensuite d'exploiter le ou les modèles les plus performants pour la génération de la cartographie, par exemple pour déterminer les zones géographiques les plus impactées par les retombées de l'incident, et pour la caractérisation du ressenti de la population, par exemple pour détecter des sentiments pré-identifiés dans tous les témoignages à disposition.

Profil du candidat :

- Étudiant en Master 2 ou en dernière année d'école d'ingénieurs, dans une spécialité de l'informatique ou des mathématiques appliquées
- Compétences requises en apprentissage automatique et en programmation (Python de préférence)

Candidature :

Les candidatures sont à envoyer à :

- Simon BERNARD, University of Rouen Normandy, simon.bernard@univ-rouen.fr
- Clément CHATELAIN, INSA Rouen Normandy, clement.chatelain@insa-rouen.fr
- Alexandre PAUCHET, INSA Rouen Normandy, alexandre.pauchet@insa-rouen.fr

Les documents à joindre à la candidature sont :

- curriculum vitae
- lettre de motivation
- résultats académiques récents
- noms et coordonnées d'un ou de plusieurs enseignants référents

⁵ Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998-6008).