

## Sujet de stage M2

### Analyse automatique des tweets pour une qualification rapide de l'intensité des crues-éclair et de leurs effets

#### Contexte :

Il est souvent difficile de percevoir rapidement l'étendue des conséquences d'une catastrophe naturelle, lorsque les informations parcellaires en provenance du terrain arrivent au compte-goutte. C'est pourtant sur la base de ce diagnostic que doit s'organiser la gestion de la catastrophe, en commençant par celle des secours, puis de l'assistance aux sinistrés.

Depuis une dizaine d'années, l'expérience montre cependant que la survenue de catastrophes naturelles se traduit fréquemment par la diffusion rapide et massive de messages sur les médias sociaux, et tout particulièrement sur Twitter dont le principe de messages courts est particulièrement bien adapté à l'expression instantanée de témoignages sur le vif. A l'instar d'autres phénomènes rapides non prévisibles tels que les séismes, les crues éclair donnent en particulier lieu à des pics d'activité très marqués sur Twitter, avec une explosion du nombre de tweets échangés après quelques minutes seulement. Cette particularité permet d'envisager le calage de modèles de détection capables de repérer automatiquement l'occurrence de ces crues ainsi que la qualification de leur impact.

En France, le BRGM a récemment mené le développement d'une plateforme participative d'analyse semi-automatique des tweets en lien avec les catastrophes naturelles, pour favoriser cette rapide remontée d'informations par les « citoyens capteurs ». Cette plateforme ([www.suricatenat.fr](http://www.suricatenat.fr)) offrira au travail de stage une infrastructure robuste de collecte de tweets liés aux crues rapides, ainsi que des fonctionnalités d'enrichissement automatique (classification à la volée de l'information, et géolocalisation).

#### Objectifs du stage :

Le stage proposé vise à définir et expérimenter une méthodologie d'analyse semi-automatique des tweets dédiée aux phénomènes de crues éclair ou crues soudaines. Cet objectif général nécessite d'aborder les questions et étapes suivantes :

1. A l'aide du consortium d'URL, établir les principales caractéristiques des crues rapides pour lesquelles des informations utiles sont à cibler dans le flux de messages échangés sur Twitter ;
2. Réaliser un état de l'art des méthodes d'analyse automatique des tweets adaptés aux crues rapides aux autres phénomènes présentant des similarités dans les phénomènes naturels à l'œuvre ou dans les cinétiques ;
3. Consolider un catalogue de crues rapides de référence survenues en France ces dernières années, sur la base duquel construire une stratégie d'analyse automatisée des Tweets (catalogue regroupant à la fois des dataset de tweets, mais également des données de référence liées à l'intensité du phénomène et à ses effets). Ce catalogue pourra s'appuyer sur les études de cas déjà identifiées dans le cadre du projet ANR PICS ([pics.ifsttar.fr](http://pics.ifsttar.fr)), et sélectionnées sur un double critère d'intensité des réactions hydrologiques constatées et d'intensité des impacts.

4. Proposer une adaptation/hybridation des méthodologies existantes (« topic modelling » par classification supervisée / semi-supervisée / non-supervisée, traitement automatique de la langue, etc.) dédiée aux phénomènes de crues rapides à étudier.
5. Réaliser une démonstration/preuve de faisabilité sur certains des datasets tweets rassemblés dans le catalogue de référence. Il s'agira notamment d'illustrer la capacité de l'information rassemblée à rendre compte de l'intensité de la crise à différentes échelles spatiales (tronçon de cours d'eau, commune, bassin versant, département, ...). La possibilité d'identifier des zones inondées à partir de cette information pourra également être évaluée.

### **Profil recherché**

Etudiant en master 2 ou école d'ingénieur, avec des compétences en programmation, machine learning, et analyse du langage.

### **Durée**

5 à 6 mois à compter de février-mars 2020.

### **Rémunération**

Gratification au taux légal de 3,75 euros de l'heure.

### **Localisation – Laboratoire d'accueil**

Ce stage est financé dans le cadre du projet pluridisciplinaire [Urba Risk Lab](#) financé par [l'I-Site FUTURE](#).

Il sera co-encadré par des chercheur.e.s issus de deux laboratoires :

- [Ifsttar, département GERS](#) (géotechnique, environnement, risques naturels et sciences de la terre)
- [BRGM, direction risques et prévention](#)

Le stage aura lieu dans les locaux du centre de Nantes de l'Ifsttar :

**Ifsttar, allée des ponts et chaussées, CS 5004, 44344 Bouguenais cedex**

### **Responsable(s) du stage et contacts**

Olivier Payrastre (Ifsttar, [olivier.payrastre@ifsttar.fr](mailto:olivier.payrastre@ifsttar.fr))

Eric Gaume (Ifsttar, [eric.gaume@ifsttar.fr](mailto:eric.gaume@ifsttar.fr))

Cécile Gracianne (BRGM, [c.gracianne@brgm.fr](mailto:c.gracianne@brgm.fr))

Faïza Boulahya (BRGM, [F.Boulahya@brgm.fr](mailto:F.Boulahya@brgm.fr))