

# Appel à candidature

## Stagiaire en Informatique pour Projet ARI-M «TS-EVAL» (action de recherche incitative orientée Master)

**Mots clés :** Apprentissage automatique, Classification supervisée, Classification non supervisée, Qualité des données, Base d'apprentissage, Corpus, Représentativité, Justesse des modèles, Métrique d'évaluation.

### Informations Supplémentaires :

- **Superviseur :** Dr Mohamed Amine BOUDIA
- **Laboratoire :** LAMIH UMR CNRS 8201 : Laboratoire d'Automatique de Mécanique et d'Informatique industrielles et Humaines
- **Département :** Informatique
- **Groupe :** Sécurité-Système-Donnée (SSD)
- **Date de début de stage :** 01/04/2024
- **Durée :** 6 mois
- **Rémunération :** 550 euro
- **Lieu de travail :** Bâtiment JONAS, campus Mont Houy, Valenciennes

### Éléments de Candidature :

- CV
- Lettre de Motivation
- Derniers Relevés de Notes Académiques

### Veillez envoyer ces documents à l'adresse suivante :

- **Objet :** [ARIM/ TS-EVAL – poste] Stage M2
- **Email :** mohamedamine.boudia@uphf.fr

Le département Informatique du laboratoire LAMIH lance un appel à candidature pour un stagiaire de master ou ingénieur en 4ème ou 5ème année en informatique pour participer à un projet de recherche dans le cadre de l'action de recherche incitative orientée Master (ARI-M) passionnant intitulé "Elaboration d'une nouvelle métrique d'évaluation et validation de qualité de la base d'apprentissage" (TS-EVAL).

### **Présentation du LAMIH :**

Le Laboratoire d'Automatique, de Mécanique et d'Informatique Industrielles et Humaines (LAMIH) est une unité mixte de recherche entre l'Université Polytechnique Hauts de France (UPHF) et le Centre National de la Recherche Scientifique (CNRS). Spécialisé dans le domaine du transport et de la mobilité humaine, le LAMIH UMR CNRS 8201 joue un rôle essentiel dans la recherche liée aux véhicules non polluants, au transport intelligent, à l'aide à la conduite, à l'écoconduite, à l'allègement des structures, à la logistique des transports, à la mobilité pour tous et à la mobilité intelligente.

Voici quelques points clés sur le LAMIH UMR CNRS 8201 :

1. **Pluridisciplinaire et reconnu** : Le LAMIH est un laboratoire pluridisciplinaire qui excelle dans les domaines du transport et de la mobilité. Il se concentre sur des sujets tels que les véhicules respectueux de l'environnement, le transport intelligent, l'aide à la conduite et la logistique des transports.
2. **Organisation en 4 départements** : Le laboratoire est structuré en quatre départements :
  - Automatique
  - Informatique
  - Mécanique
  - Sciences de l'Homme et du Vivant (SHV)
3. **Membre de la Fédération de Recherche du CNRS Transports Terrestres & Mobilité (FR 3733)** : Le LAMIH UMR CNRS 8201 est un acteur clé de la recherche dans ce domaine et co-dirige la Fédération de Recherche du CNRS Transports Terrestres & Mobilité.
4. **Partenaire du développement de l'IRT Railenium** : En plus de sa collaboration avec le CNRS, le LAMIH est un partenaire privilégié du développement de l'IRT Railenium, contribuant ainsi à l'innovation dans le secteur ferroviaire.
5. **Label Carnot et membre de l'Institut Carnot ARTS** : Grâce à sa recherche partenariale d'excellence, le LAMIH UMR CNRS 8201 a obtenu le label Carnot et est membre de l'Institut Carnot ARTS.

### **Description du projet:**

Le projet TS-EVAL vise à élaborer une nouvelle métrique d'évaluation de la qualité de la base d'apprentissage en se basant sur des critères théoriques pour la génération d'un training set optimal. L'objectif principal est de renforcer les évaluations existantes en mettant l'accent sur la représentativité de la base d'apprentissage par rapport au corpus, afin d'améliorer la justesse des processus de classification et la fiabilité des modèles d'apprentissage générés. Le stagiaire participera à toutes les étapes du projet, y compris la recherche bibliographique, le développement de la métrique proposée, les expérimentations et l'analyse des résultats.

Voici les points clés du projet :

- Contexte : Le projet se situe dans le domaine de l'apprentissage automatique et de la classification. Il s'intéresse spécifiquement à la qualité des données utilisées pour entraîner les modèles.
- Objectif principal : L'objectif central est d'améliorer les évaluations existantes en mettant l'accent sur la représentativité de la base d'apprentissage par rapport au corpus global. En d'autres termes, il s'agit de s'assurer que les données d'entraînement reflètent fidèlement la diversité et les caractéristiques du domaine d'intérêt.
- Justesse des processus de classification : En optimisant la qualité de la base d'apprentissage, on vise à améliorer la précision des modèles de classification. Des données d'entraînement plus pertinentes et représentatives conduiront à des prédictions plus fiables.

### **Tâches du stagiaire :**

1. Effectuer une revue de la littérature sur les méthodes d'évaluation de la qualité de la base d'apprentissage.
2. Contribuer à la conception et au développement de la nouvelle métrique d'évaluation TSEVAL.
3. Mettre en œuvre des expériences pour évaluer la performance de la métrique proposée sur des ensembles de données réels.
4. Analyser les résultats expérimentaux et rédiger des rapports et des articles scientifiques.
5. Participer à des réunions régulières avec l'équipe de recherche pour discuter des progrès et des résultats du projet.

### **Profil du candidat :**

- Étudiant(e) en cours de validation d'un Master en Informatique, Data Science ou domaine connexe, ou étudiant de 4<sup>ième</sup> ou 5<sup>ième</sup> année Ingénieur en Informatique.
- Solides compétences en programmation (Java, Python, R, etc.) et en traitement des données.
- Bonne compréhension des concepts d'apprentissage automatique et de validation de modèles.
- Capacité à travailler de manière autonome et à résoudre des problèmes de manière créative.
- Bonnes compétences en communication écrite et orale.

### **Durée et conditions du stage :**

- Début de stage : à partir de 01/04/2024
- Durée : 6 mois.
- Rémunération : 550 euros par mois.
- Le stagiaire bénéficiera d'un encadrement étroit de l'équipe de recherche.
- Possibilité de participer à 1 ou 2 conférences national ou internationales en fonction des résultats.

### **Comment postuler :**

Les candidat(e)s intéressé(e)s sont invité(e)s à soumettre leur candidature par e-mail à [mohamedamine.boudia@uphf.fr](mailto:mohamedamine.boudia@uphf.fr) avant le 21/03/2024 . La candidature devrait inclure :

1. Un CV détaillé comprenant les informations sur l'éducation, l'expérience de recherche et les compétences en programmation.
2. Une lettre de motivation décrivant l'intérêt du candidat pour le projet de recherche TSEVAL et expliquant comment ses compétences et son expérience répondent aux exigences du poste.
3. Derniers Relevés de Notes Académiques
4. Les coordonnées de deux références académiques.

Nous sommes impatients de recevoir votre candidature et de vous accueillir au sein de notre équipe de recherche pour contribuer à ce projet passionnant.

### **Bibliographie :**

1. Twomey, J. M., & Smith, A. E. (1997). Validation and verification. Artificial neural networks for civil engineers: Fundamentals and applications, 44-64.
2. Cheng, J., Xu, Z., Wu, W., Zhao, L., Li, X., Liu, Y., & Tao, S. (2014). Training set selection for the prediction of essential genes. PloS one, 9(1), e86805.
3. Nalepa, J., & Kawulok, M. (2019). Selecting training sets for support vector machines: a review. Artificial Intelligence Review, 52(2), 857-900.
4. Ryan, K. N., Bahhur, B. N., Jeiran, M., & Vogel, B. I. (2021, April). Evaluation of augmented training datasets. In Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXXII (Vol. 11740, pp. 118-125). SPIE.
5. Boudia, M. A. (2023, October). Evaluate the Training Set: is it necessary? A theorecal Presentation of Y\_Measure: New Metric for Evaluating Training Sets for Supervised Classification. In 2023 IEEE International Conference on Networking, Sensing and Control (ICNSC) (Vol. 1, pp. 1-6). IEEE.

### **Mohamed Amine BOUDIA**

Maître de conférences

Membre du LAMIH - UMR CNRS 8201 Laboratoire d'Automatique de Mécanique et d'Informatique industrielles et Humaines

Département informatique, INSA Haut-de-France (Institut National des Sciences Appliquées) Université Polytechnique Hauts-de-France

Campus du Mont Houy, bâtiment Jonas - Bureau 212A59313  
Valenciennes Cedex 9, France

Tél : +33 3 27 51 19 44

Mobile : +33 7 45 68 77 89

[consultez l'accessibilité de notre bâtiment](#)