

# Stage de Master 2 (5-6 mois)

**Lieu :** Laboratoire LIRIS (Lyon, France)/ Entreprise Hopliz

**Responsables du stage (Emails) :**

- Hamida Seba ([hamida.seba@univ-lyon1.fr](mailto:hamida.seba@univ-lyon1.fr))
- Philippe Manipoud ([philippe.manipoud@hopliz.com](mailto:philippe.manipoud@hopliz.com))

**Thématiques :** Apprentissage automatique, profilage de clients, détection d'anomalies, graphes/graphes de connaissances, données hétérogènes

**Type de stage :** Fin d'études bac +5, Master 2

**Durée :** 5-6 mois

**Période souhaitée :** idéalement à partir de février 2024

**Rémunération :** gratification de stage

**Intitulé :** Modèles d'apprentissage automatique sur des données complexes pour la prévention de défauts de paiement

**Sujet :**

Pour plusieurs applications du monde réel, identifier des motifs qui ne se conforment pas à une activité normale est une question fondamentale pour garantir une prestation de service correcte ainsi que la sécurité et la fiabilité du système [1]. C'est principalement le cas pour des applications de surveillance et de suivi telles que la vidéosurveillance, la surveillance médicale, la détection de logiciels malveillants, la détection de fraudes financières, etc. Un motif anormal est appelé anomalie ou valeur aberrante. Une anomalie est généralement définie comme un motif comportemental qui s'écarte significativement de la plupart des motifs comportementaux du système surveillé et apparaît dans une proportion significativement plus petite que celle des motifs normaux. Avec l'explosion des quantités de données à traiter pour ce genre d'applications, le recours aux modèles d'apprentissage et en particulier l'apprentissage profond est devenu inévitable dans ce domaine.

Durant ce stage, nous nous intéressons aux modèles d'apprentissage capables de traiter des données complexes multi-sources et hétérogènes sur les clients d'une entreprise de location d'ordinateurs personnels (MacBook & iPad) :

- données contrôlées communiqués par le client (nom, prénom, adresses (facturation, livraison), tel, mail)
- données informatique extraites de la connexion internet du client (IP)
- données bancaires
- données issues de blacklists
- données socio-démographiques issues de l'INSEE

Le but de l'analyse de ces données est de réaliser un profiling plus « fair » (sans critères socio-discriminants) de clients qui permettra de prévenir les défauts de paiement. L'approche proposée est la construction de graphes de connaissances ego-centriques [2] profilant les clients et permettant de représenter toutes les informations les concernant. Ensuite, il s'agira de se baser sur cette représentation pour détecter toute anomalie qui peut engendrer un défaut de paiement. Cette représentation basée sur des graphes de connaissances nécessitera d'utiliser et/ou de concevoir des modèles d'apprentissage adaptés [3, 4]. Le stagiaire commencera par effectuer un état de l'art sur la problématique ainsi que sur les modèles d'apprentissage existants. Il mettra ensuite en place la solution la plus adaptée aux données considérées durant le stage.

**Ce stage pourra se poursuivre avec une thèse CIFRE.**

**Profil du/de la stagiaire :** Compétences avancées (niveau M2) en informatique (en particulier en apprentissage machine fortement souhaitées).

- **Merci d'adresser votre candidature avec un CV ainsi que vos notes de l'année universitaire en cours et de l'année 2022/2023 à [hamida.seba@univ-lyon1.fr](mailto:hamida.seba@univ-lyon1.fr) et [philippe.manipoud@hopliz.com](mailto:philippe.manipoud@hopliz.com)**

## Références

- [1] Abd Errahmane Kiouche, Sofiane Lagraa, Karima Amrouche & Hamida Seba (2021). "A simple graph embedding for anomaly detection in a stream of heterogeneous labeled graphs". Pattern Recognition, p. 107746. doi : [10.1016/j.patcog.2020.107746](https://doi.org/10.1016/j.patcog.2020.107746). HAL : [hal-02993787](https://hal.archives-ouvertes.fr/hal-02993787)
- [2] Farouk Damoun, Hamida Seba, Jean Hilger & Radu State (2023). "G-HIN2Vec: Distributed heterogeneous graph representations for cardholder transactions". SAC '23: 38th ACM/SIGAPP Symposium on Applied Computing, 31 mars 2023, Tallinn Estonia (France), pp. 528-535. doi : [10.1145/3555776.3577740](https://doi.org/10.1145/3555776.3577740). HAL : [hal-04238052](https://hal.archives-ouvertes.fr/hal-04238052)
- [6] William L. Hamilton. Graph Representation Learning. E-bbok Springer
- [3] Carl Yang, Yuxin Xiao, Yu Zhang, Yizhou Sun, and Jiawei Han. 2020. Heterogeneous network representation learning: A unified framework with survey and benchmark. IEEE Transactions on Knowledge and Data Engineering (2020).
- [4] Susie Xi Rao, Shuai Zhang, Zhichao Han, Zitao Zhang, Wei Min, Zhiyao Chen, Yinan Shan, Yang Zhao, and Ce Zhang. 2020. xFraud: Explainable fraud transaction detection on heterogeneous graphs. arXiv preprint arXiv:2011.12193 (2020).