

## Post-doctorat en sciences de données : graphes probabilistes et classification appliqués à l'analyse des défauts en production micro-électronique

**Type de poste :** Post-doctorat (candidat disposant d'une thèse)

**Contrat :** Poste ouvert en CDD.

**Durée :** 24 mois

**Responsable Scientifique :** Pr. Rodolphe Le Riche, CNRS senior researcher, EMSE/FAYOL - LIMOS.

**Responsable du Projet :** Pr. Xavier Boucher, EMSE – Institut FAYOL - LIMOS.

**Date de démarrage proposée :** Octobre 2020

**Collaboration industrielle et européenne :** STMicroelectronics (Grenoble Reliability & Analysis Laboratory), Projet Européen FA4.0 incluant 15 partenaires européens scientifiques et industriels.

**Mots clés :** Data-mining, Machine Learning, Process-mining, Classification, Artificial Intelligence, Industrie micro-électronique

### Contexte scientifique et industriel

Ce post-doctorat s'inscrit dans le cadre du projet européen FA4.0, en collaboration avec les équipes de STMicroelectronics sur Grenoble (Grenoble Reliability & Analysis Laboratory). Plus précisément, le projet vise à créer et déployer des solutions d'Intelligence Artificielle développées pour transformer et améliorer les pratiques industrielles au sein d'un service particulier de l'entreprise : le laboratoire d'analyse des défaillances. Ce laboratoire est un maillon essentiel de la production industrielle, en charge d'analyser et de diagnostiquer les problèmes de qualité et de défaillance qui apparaissent dans les processus de production de wafers micro-électroniques (fabrication de micro-puces) et de produits complexes.

L'analyse des défaillances tout au long de la chaîne de valeur à partir d'une puce, est une condition préalable à l'amélioration de la fiabilité et de la qualité et donc de la compétitivité des dispositifs électroniques, en particulier sur le marché de l'automobile et de l'industrie pour les applications exigeantes en matière de fiabilité et de sécurité. La forte progression de la numérisation et de l'automatisation offre un formidable potentiel pour mieux maîtriser la fiabilité des produits de haute technologie qui sont basés sur des systèmes électroniques de plus en plus complexes.

Grâce à l'apport du digital, l'objectif du projet est donc de transformer les pratiques actuelles du laboratoire d'analyse de défaillance, en mobilisant les techniques du machine learning. Ces techniques permettent d'enrichir la capacité explicative des modèles statistiques, par une capacité renforcée à discriminer les causes possibles de non fiabilité et à prendre en compte leurs interactions. Plus généralement, le projet s'inscrit dans les travaux sur l'efficacité du diagnostic par apprentissage, en particulier par classification.

### Problématique, Objectifs et Mission de recherche

La spécificité d'un laboratoire d'analyse des défaillances est d'opérer un certain nombre de tests électriques, physiques et chimiques, sur les supports micro-électroniques qui contiennent les composants électroniques susceptibles d'être atteint par un défaut (défaillance). Lors de cette analyse, la succession des tâches d'exploration dépend des résultats observés jusqu'à l'étape  $t$  ; la décision pour le choix de l'étape de test suivant  $t+1$  dépend des résultats obtenus avant la phase  $t$ , et de l'expertise humaine. Ce diagnostic, appliqué à un support électronique sur lequel un défaut a été observé, peut être vu comme un parcours sur un graphe de décision où chaque nœud est un choix parmi un ensemble d'alternatives possibles (alternatives : quelle est l'étape d'analyse suivante à réaliser ?). Les résultats des analyses et ou des tests se présentent sous la forme de données variées (notamment d'importantes bases de données d'imagerie par microscopies diverses, de signaux physiques, de mesures, de données textuelles ...) qui sont utilisées par l'analyste pour la construction de son diagnostic au

cours du processus d'analyse du défaut.

En collaboration avec un doctorant et avec l'équipe de recherche, le post-doctorant sera en situation d'assurer le pilotage opérationnel du déroulement des tâches scientifiques à réaliser durant le projet. Le post-doctorant prendra lui-même en charge les travaux relevant de l'apprentissage supervisé, tout en assurant la cohérence et la coordination de l'ensemble des avancées scientifiques. Les travaux scientifiques à développer viseront à la fois des avancées scientifiques et théoriques et la réalisation de preuves de concepts démontrées à partir de données réalistes issues du milieu industriel, dans le but de favoriser l'intégration ultérieure des résultats par les industriels.

Les méthodes anticipées pour réaliser ces travaux sont les modèles de graphes probabilistes (réseaux bayésiens, modèles de Markov) et leur utilisation au sein de procédures de classification supervisée.

Exemples de question scientifiques :

- Analyse de la structure des données/information/expertise à considérer et spécification de l'ensemble des dimensions et variables probabilistiques à considérer pour l'apprentissage
- Gestion de l'hétérogénéité des données/informations dans les processus d'apprentissage.
- Nécessité de créer une topologie des 'Hypothèses explicatives' actuellement disponibles dans la base de données.
- Recherche de parcours optimaux dans un graphe à des fins de classification.

Exemples de questions industrielles :

- Dans quelles situations le 'Failure analysis flow' peut-il être simplifié ?
- Grâce aux connaissances extraites par apprentissage, peut-on anticiper la construction des 'Hypothèses explicatives' en supprimant ou réduisant certaines étapes d'analyse ?
- Peut-on identifier des parcours optimaux pour la totalité des 'parcours décisionnels' ?
- Comment gérer l'hétérogénéité des données ?
- Quelles préconisations doivent être prises en compte pour une meilleure structuration des données pour l'avenir ?

## Profil du candidat

Le candidat en post-doctorat devra disposer des compétences initiales de niveau ingénieur/M2 dans le domaine des mathématiques appliquées, avec une orientation vers les sciences des données, les modèles probabilistes, le machine learning. Il devra avoir soutenu sa thèse au moment du démarrage de poste.

Un parcours de recherche et des expériences dans un ou plusieurs des domaines suivants seront appréciées :

- Mathématiques appliquées, orientées vers l'analyse et le traitement des données ;
- Science des données, Machine learning, Process mining, Intelligence artificielle, identification de modèles.
- Apprentissage supervisé, Classification, Théorie des graphes
- Motivation pour l'application des méthodes développées, notamment pour le diagnostic en contexte industriel.

Bien sûr un intérêt ou des expériences complémentaires en lien avec l'industrie de la microélectronique seront également appréciées.

Une capacité à publier en revue scientifique, ainsi qu'un bon niveau d'anglais sont attendus.

## Procédure de candidature

Les dossiers de candidature devront comprendre :

- Un curriculum vitae faisant état des activités d'enseignement, des travaux de recherche et, éventuellement, des relations avec le monde économique et industriel,
- Lettre de Motivation
- A la discrétion des candidats, des lettres de recommandation,
- La copie du doctorat (ou PhD),
- La copie d'une pièce d'identité

Elles doivent être adressées sous forme électronique avant le 30 Juin 2020 à :

Rodolphe LE RICHE  
Ecole Nationale Supérieure des Mines de Saint-Étienne  
158 cours Fauriel, CS62363, 42023 Saint-Étienne Cedex 2  
Tel: (33) (0) 477420023  
Email: [leriche@emse.fr](mailto:leriche@emse.fr)  
Copie à Xavier BOUCHER : [boucher@emse.fr](mailto:boucher@emse.fr)