

## Fouille de données condensée à partir de séries temporelles pour une prédiction explicable

**Financement** : projet ANR XQUALITY

**Superviseurs post-doctorants** : Ahmed SAMET

**Adresse mail** : [ahmed.samet@insa-strasbourg.fr](mailto:ahmed.samet@insa-strasbourg.fr)

**Lieu de travail** : INSA Strasbourg

**Laboratoire** : Laboratoire ICUBE

**Équipe**: Science de données et connaissances

**Démarrage prévu** : La date de départ pourrait être planifiée avec le candidat sélectionné après le 1er janvier 2023

### Le projet en quelques mots :

Les entreprises française et allemandes sont connues pour leur production de haute qualité et leur orientation vers les usines intelligentes. Le contrôle de la qualité des systèmes de production complexes constitue un défi majeur et sont encore exacerbés par la pénurie de main-d'œuvre qualifiée dans ce domaine. Les problèmes de qualité, constaté dans une ligne de production, doivent être détectés et éliminés rapidement. Lorsqu'un problème qualité est détecté, il est nécessaire d'en comprendre rapidement les multiples causes possibles (qui peuvent parfois être contradictoires) afin de proposer les actions correctives les plus appropriées pour ramener le processus de fabrication à son mode de fonctionnement normal.

Le projet **XQuality** étudie des approches d'IA hybrides et explicables pour aider les entreprises à mettre en œuvre un contrôle qualité intelligent et automatisé. Le projet combine l'apprentissage automatique, les technologies sémantiques et les connaissances d'experts pour surveiller et expliquer la perte de qualité des produits et des processus dans une entreprise. L'objectif est de développer un système basé sur l'IA qui aidera le personnel à identifier les principales causes des problèmes de qualité le plus tôt possible, afin de parvenir à une ingénierie de fiabilité (reliability engineering en anglais) dans le domaine de la fabrication, grâce aux nouveaux modèles d'assurance qualité.

### Sujet du postdoctorat:

**XQuality** développera de nouveaux algorithmes de fouille de données qui analyseront les données de séries temporelles et généreront un ensemble condensé de motifs riches appelés chroniques (Sellami et al. 2020). Malheureusement, comme le souligne la littérature, le nombre de chroniques est encore élevé et leur extraction coûte beaucoup de temps. Par conséquent, nous visons à réduire l'ensemble des modèles générés en optimisant plusieurs mesures de qualité telles que la discrimination, la confiance pour n'en citer que quelques-unes (Sahoo et al 2015) visant à des modèles plus explicables.

D'un point de vue méthodologique, nous visons à proposer un nouvel algorithme de fouille de données pour extraire des motifs riches et séquentiels appelés chroniques. Un nouvel ensemble condensé basé sur la définition de fermeture des motifs est proposé. De plus, les documents de dépannage (troubleshooting en anglais) des machines seront analysés à l'aide d'une technique basée sur le CRF

pour trouver des actions correctives à chaque situation de perte de qualité. L'idée est d'associer des modèles à des situations et à des actions correctives. Cette qualité de cet ensemble est évaluée sur une tâche prédictive pour évaluer la précision de la prédiction des pertes de qualité. Nous prolongeons le travail de (Sellami et al 2020) en intégrant plusieurs contraintes dans l'algorithme de fouille de chroniques. Ces contraintes de qualité sont liées soit à la connaissance du domaine, soit à des mesures de règles de qualité telles que la discrimination et la confiance.

Deux directions pourraient être entreprises : soit par programmation déclarative, soit par un processus d'optimisation (Guyet et al 2017). L'ensemble condensé de chroniques trouvé est appliqué à une tâche de prédiction de la perte de qualité pour assurer la surveillance des pertes de qualité.

**Mots-clés :** fouille de motifs, ensembles condensés, Programmation par contraintes, IA explicable, Modélisation de l'incertitude, Industrie 4.0, Contrôle qualité.

**Compétences du candidat :** Le travail exige une formation solide en fouille de données et en mathématiques. Une expérience préalable dans la fouille de données, de la programmation par contraintes et de la modélisation de l'incertitude est fortement recommandée. Avoir une contribution au doctorat en fouille de données et en IA explicable sur des séries temporelles est très apprécié. De plus, l'expérience antérieure sur l'industrie 4.0 et le contrôle de la qualité est très appréciée.

**Connaissances spécifiques :** Une connaissance en fouille de motifs dans les séries temporelles est attendue du candidat(e). Les candidat(es) doivent démontrer que leur profil correspond à la description du projet. Avoir publié dans des conférences de fouille de données de premier plan est très apprécié.

**Formation souhaitée:** Le profil de doctorat junior en informatique est souhaité.

**Salaire :** Salaire brut mensuel (HT) : ~ 2500 euros et selon l'échelle salariale de l'INSA

**Durée du poste:** 1 an. Pourrait être prolongé en soumettant à des appels externes.

### **References:**

Sellami, C., Miranda, C., Samet, A., Tobji, M. A. B., & de Beuvron, F. (2020). On mining frequent chronicles for machine failure prediction. *Journal of Intelligent Manufacturing*, 31(4), 1019-1035.

Guyet, T., Happe, A., & Dauxais, Y. (2017, June). Declarative sequential pattern mining of care pathways. In *Conference on Artificial Intelligence in Medicine in Europe* (pp. 261-266). Springer, Cham.

Sahoo, J., Das, A. K., & Goswami, A. (2015). An effective association rule mining scheme using a new generic basis. *Knowledge and Information Systems*, 43(1), 127-156.