

# Découverte de sous-groupes de prédictions interprétables pour le triage d'incidents

Youcef Remil<sup>\*,\*\*\*</sup>, Anes Bendimerad<sup>\*\*\*</sup>, Marc Plantevit<sup>\*\*</sup>  
Céline Robardet<sup>\*</sup>, Mehdi Kaytoue<sup>\*,\*\*\*</sup>

<sup>\*</sup>Université de Lyon, CNRS, INSA-Lyon, LIRIS, UMR5205, F-69621, France

<sup>\*\*</sup>EPITA Research and Development Laboratory (LRDE), France

<sup>\*\*\*</sup>Infologic, 99 avenue de Lyon, F-26500, Bourg-Lès-Valence, France  
yre@infologic.fr

**Résumé.** Le besoin de maintenance prédictive s'accompagne d'un nombre croissant d'incidents qui doivent être rapidement assignés aux services appropriés pour des actions correctives. Il existe des modèles prédictifs pour automatiser cette assignation, mais les plus efficaces sont opaques. Des méthodes ont été conçues pour expliquer localement chaque prédiction de tels modèles, mais elles fournissent une explication à chaque résultat, inconcevable en présence d'un nombre important de prédictions à analyser. Nous proposons d'abord un modèle efficace de triage d'incidents, puis une méthode basée sur la découverte de sous-groupes pour grouper les explications de ses prédictions. Cette méthode permet (1) de grouper les incidents dont les prédictions partagent des explications similaires et (2) de fournir une description interprétable à chacun de ces sous-groupes d'incidents. Cet article est une traduction résumée de (Remil et al., 2021).

## 1 Introduction

De nos jours, de nouvelles technologies sont constamment introduites et de nouvelles perspectives de numérisation sont adoptées en industrie : des logiciels comme l'Enterprise Resource Planning (ERP) sont au centre de cette révolution. Il s'ensuit que la maintenance de ces systèmes logiciels est devenue de plus en plus complexe car impliquant toujours plus de composants physiques, logiciels et métiers interdépendants. En effet, les incidents sont de diverses natures : (i) un incident prédit à partir des modèles construits sur des données historiques, (ii) un incident réel non encore signalé par l'utilisateur final mais par le système de supervision et (iii) finalement, un incident rapporté par l'utilisateur final, mais généralement mal décrit et contextualisé. Lorsque le nombre d'incidents remontés est important, il devient difficile pour les ingénieurs d'astreinte de les évaluer rapidement et les orienter vers les services appropriés pour mener des actions correctives. Afin d'accélérer et d'améliorer ces décisions, plusieurs modèles prédictifs ont été proposés pour le *trriage d'incidents*, dont le but est de prédire efficacement pour un incident signalé le service à cibler. Les meilleures approches de l'état de l'art utilisent des modèles d'apprentissage profond (Chen et al., 2019; Pham et al., 2020; Wang et al., 2021). Toutefois, ces modèles sont opaques et difficiles à interpréter puisqu'ils ne révèlent pas la logique derrière leur mécanisme de prise de décision. Un défi important pour