SEDAF : Prototype d'un Système Explicable de Détection d'Anomalies dans les Flux de Données

F. Jiechieu Kameni, A. M. S. Ngo Bibinbe, V. Cako, A. J. Djiberou Mahamadou, M. R. Bakari, K. D. Nguetche, D. Kamga Nguifo, A. Bertrand, M. F. Mbouopda, R. El Cheikh, G. R. Mbiadou Saleu, E. Mephu Nguifo

Université Clermont Auvergne, Clermont Auvergne INP, ENSMSE, CNRS, LIMOS, F-63000 Clermont–Ferrand, France

Résumé. La détection d'anomalies fait référence à l'identification des évènements rares qui diffèrent grandement de la tendance normale et majoritaire observée dans la distribution des données. Lorsque le nombre de variables à analyser est important, il peut être difficile de comprendre l'anomalie détectée sans explication. Dans ce travail, nous présentons le prototype d'un système explicable de monitoring et de détection d'anomalies en temps réel, à partir de mesures provenant d'un flux de données. Le système construit est composé d'une combinaison de méthodes de détection d'anomalies alliant apprentissage profond et arbres de décision ainsi que d'une méthode d'explicabilité agnostique. Dans un contexte d'apprentissage non supervisé, nous montrons également comment l'explicabilité fournit des éléments de validation du système en combinaison avec les retours des experts du domaine.

1 Introduction

Les flux de données font référence à des données massives qui arrivent en continue et en temps réel. Un flux de données est en théorie infini. On retrouve très souvent les flux de données dans plusieurs contextes : ils peuvent être générés par des capteurs, les réseaux sociaux ou par les utilisateurs d'une application. La détection des anomalies dans les flux de données permet d'identifier des irrégularités dans la distribution du flux de données. Ses applications sont multiples : dans le domaine de l'agriculture, l'analyse des flux de données générés par les capteurs placés au sol permet de surveiller la composition du sol et garantir un bon rendement des récoltes. Dans la manufacture des composants électroniques, l'analyse de l'air ambiant permet de contrôler la contamination et de prévenir la dégradation des composants électroniques. Les contraintes de ces environnements nécessitent souvent de faire une analyse en temps réel car, passé un certain delai, les dégats peuvent déjà être considérables en cas d'anomalie.

Plusieurs méthodes de détection d'anomalies (Mirsky et al. (2018); Arslan et al. (2023); Tan et al. (2011); Mensi et Bicego (2021)) ont été étudiées dans le cadre de ce travail. La plupart de ces méthodes a été conçue pour les séries temporelles et sont utilisées pour traiter les flux de données grace à la technique de fenêtrage (Mansalis et al. (2018)).

Dans cet article, nous décrivons la conception d'un système de détection d'anomalies et d'explication en temps réel de ces anomalies.

Le système proposé comporte 4 modules principaux :
— une base de données temporelle ;
— un module de détection d'anomalies ;
— un module d'explicabilité ;
— un module de visualisation.

Le module de détection d'anomalies s'appuie sur plusieurs approches incluant les arbres de décisions et l'apprentissage profond ; tandis que le module d'explicabilité met en oeuvre une méthode d'explication agnostique et par attribution des scores.

La principale innovation dans ce travail est la mise en oeuvre d'un système de surveillance qui permet la détection des anomalies et l'explication de ces anomalies en temps réel avec une interface utilisateur permettant de visualiser conjointement le flux, les anomalies et les explications en fonction du temps. Cette vue utilisateur permet d'analyser et de comprendre rapidement les résultats fournis par le système.

Cet article sera organisé comme suit : nous allons d'abord présenter les méthodes de détection d'anomalies, ensuite nous présenterons les méthodes d'explicabilité locales avant de présenter l'architetcure du système conçu avec ses différents modules fonctionnels ainsi que les modèles scientifiques mis à contribution. Nous terminerons par l'évaluation qualitative du système, une conclusion et des perspectives.

2 Fondements du système mis en oeuvre

Le système décrit dans cet article repose sur les concepts fondamentaux intégrant flux de données, détection d'anomalies et explicabilité.

2.1 Détection des anomalies dans les flux de données

En dehors de l'aspect temps réel, un flux de données peut être caractérisé comme une série temporelle, à savoir une succession de valeurs indexées par le temps. Plusieurs méthodes de détection d'anomalies applicables aux flux de données ont été étudiées dans le cadre de ce travail. Les résultats de cette étude sont reportés dans l'article Ngo Bibinbe et al. (2022b).

Nous avons ainsi les méthodes basées sur les arbres de décision telles que HStree (Tan et al. (2011)) ou iForest (Mensi et Bicego (2021)). Le principe général de ces méthodes est de construire un prédicteur ensembliste constitué de plusieurs arbres d'isolation. Chaque noeud de l'arbre représente un test sur une dimension. Une donnée qui est isolée très tôt dans l'arbre, c'est-à-dire, rangée dans un noeud de faible masse (contenant très peu de données) et de faible profondeur par rapport à la hauteur totale de l'arbre sera considérée comme une anomalie. Les prédictions de tous les arbres de la forêt sont agrégées pour avoir la prédiction globale.

Une seconde catégorie regroupe les méthodes basées sur l'apprentissage profond telles que KitNet (Mirsky et al. (2018)) ou DeepAnt (Arslan et al. (2023)). Les approches non supervisées utilisant les techniques d'apprentissage profond pour la détection d'anomalies dans les séries temporelles se basent pour la plupart sur des modèles génératifs ou encore des autoencodeurs (Kingma et Welling (2019)). Le principe général de ces méthodes est d'apprendre la distribution du flux, et d'utiliser le modèle appris pour reconstituer les données de la série temporelle ou du flux. Si l'erreur de reconstruction de la donnée est élevée (dépasse un certain seuil), alors, la donnée est considérée comme anormale.

Une troisième catégorie de méthodes concerne celles basées sur le clustering. La détection d'anomalies par clustering consiste à partitionner les données en formant des groupes suivant le principe de la maximisation de la similarité intra-groupe et de la minimisation de la similarité inter-groupe. Une fois les groupes constitués, les données qui sont isolées seules ou en très petit nombre seront considérées comme de potentielles anomalies. Dans cette catégorie, on peut citer Drag-Stream (Ngo Bibinbe et al. (2022a)) pour les flux univariés.

Enfin, on peut aussi concevoir des approches ensemblistes qui utilisent une combinaison d'un ou de plusieurs modèles de l'une des approches précédentes. Chaque modèle de base calcule sa décision ou son score d'anomalie, et la décision finale est obtenue en agrégeant les prédictions des modèles de base.

Toutes ces approches sont applicables et fonctionnent bien sur les séries temporelles. Mais dans un contexte du flux de données, des contraintes de temps réel sont applicables : il s'agit non seulement des contraintes liées au delai de détection des anomalies (et par conséquent de production des explications), mais aussi de la nécessité de prendre en compte des changements dans la dynamique du flux avec le temps.

2.2 Explicabilité

Lorsqu'on détecte une anomalie, dans la distribution du flux de données, il est important de fournir à l'expert du domaine une explication qui lui permettra de rapidement comprendre la nature de l'anomalie et savoir à quel niveau il pourrait intervenir. L'explication peut également permettre à l'expert d'apprécier le niveau d'alerte étant donné que ces modèles produisent très souvent de fausses alertes.

En contexte multivarié, la nécessité d'expliquer est d'autant plus prononcée qu'elle va permettre de savoir sur quelles variables s'est portée l'anomalie. Il existe plusieurs approches d'explication. Dans ce travail, nous allons nous intéresser aux explications locales par attribution de scores : il s'agit d'une classe de méthodes d'explication qui permettent d'expliquer la décision du modèle sur une instance composée de plusieurs variables, en calculant les contributions de chaque variable à la sortie du modèle (Groen et al. (2022)).

Dans cette catégorie de méthodes, nous avons les méthodes dites agnostiques (indépendantes du modèle) et les méthodes dépendantes du modèle. Les méthodes indépendantes du modèle permettent d'expliquer tout type de modèle en considérant ceux-ci comme des boîtes noires. Parmi ces méthodes nous avons la méthode SHAP (Lundberg et Lee (2017)) et LIME (Ribeiro et al. (2016)) dont le principe est d'expliquer le modèle en calculant une fonction explicable qui approxime le modèle au voisinage de l'instance de donnée pour laquelle on veut produire une explication. A l'inverse, les méthodes dépendantes du modèle à l'instar de LRP (Binder et al. (2016)) essaient d'analyser le fonctionnement interne d'un modèle au moment de la prédiction afin de déterminer les contributions des variables d'entrée par rapport à la valeur prédite par le modèle.

Ayant présenté les concepts qui encadrent ce travail, il est question à présent de décrire le système, ses principaux composants ainsi que les modèles mis à contribution pour son fonctionnement.

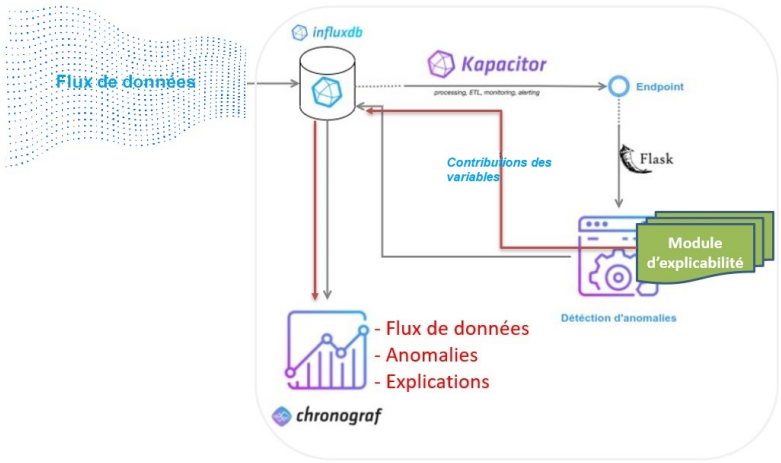


FIG. 1 – Architecture du système mis en place.

3 Architecture du système

Les principaux composants fonctionnels du système sont les suivants :
— un module de stockage de données ;
— un module d'analyse et de détection d'anomalies ;
— un module d'explicabilité en temps réel ;
— un module de visualisation.

3.1 Module de stockage de données :

Le module de stockage de données ici est constitué d'une base de données de série temporelle qui va stocker les données provenant du flux en les indexant par le temps afin d'en faciliter l'analyse. Chaque donnée du flux est constituée des valeurs de mesure (numériques) de plusieurs variables avec une date représentant l'instant de mesure de ces variables. Le choix de la base de données s'est porté sur *Influxdb* car, elle est open source et adaptée pour les applications de monitoring. Ses performances sont évaluées dans Khelifati et al. (2023).

Influxdb, *Kapacitor* qui joue le rôle d'intégrateur de données et *Chronograph* qui est un outil de visualisation font tous partie d'un même écosystème d'outils de stockage et d'analyse des séries temporelles développés par l'entreprise InfluxData [1].

3.2 Module d'analyse et de détection d'anomalies

Le module de détection des anomalies a été conçu dans l'optique de pouvoir être mis à jour au fil du temps de manière à prendre en compte les changements observés dans la dynamique

1. Entreprise de logiciels basée à San Francisco, Californie aux Etats-Unis.

des données qui arrivent continuellement. Pour traiter le flux de données qui arrive en continue et de manière ininterrompue, une technique de fenêtrage (Mansalis et al. (2018)) a été mise sur pied avec la définition d'une taille de fenêtre. La taille de la fenêtre est un paramètre qui détermine les points qui seront utilisés pour entraîner le modèle qui va servir à détecter les anomalies. Lorsque les points arrivent et qu'un nombre de points équivalent à la taille de fenêtre est traité, le modèle est mis à jour et ainsi de suite. Avec cette technique, il est possible de mettre à jour le modèle après chaque taille de fenêtre avec les données les plus récentes.

Par ailleurs, lorsque une fenêtre est remplie [2], les paramètres de normalisation z-score à savoir la moyenne et l'écart-type sont calculés sur les données de la fenêtre et sont utilisés pour normaliser les prochaines données qui arrivent jusqu'à ce que la taille de la prochaine fenêtre soit atteinte et ainsi de suite.

L'un des objectifs visés dans le cadre de ce projet était de founir aux utilisateurs du système, une variété d'algorithmes de détection d'anomalies paramétrables en fonction des spécificités du problème. C'est dans ce sens qu'un ensemble d'algorithmes de détection d'anomalies a été préliminairement étudiés et intégrés. Les résultats de ces études sont répertoriés dans (Ngo Bibinbe et al. (2022b)). Il s'agit de : HS-Tree et IForest pour les arbres de décision, DeepAnt et KitNet pour l'apprentissage profond et d'un modèle ensembliste basé sur le principe du vote majoritaire.

3.3 Module d'explicabilité

Chaque algorithme de détection d'anomalies retourne un score d'anomalie, et un algorithme d'explication est exécuté pour expliquer le score d'anomalie d'une mesure à un instant précis, lorsque ce score est supérieur à une valeur seuil. La méthode d'explication mise en œuvre est la méthode SHAP en particulier KernelShap (Lundberg et Lee (2017)). L'intérêt d'une méthode agnostisque comme KernelShap se justifie principalement par le fait que plusieurs modèles de détection d'anomalie ont été intégrés et SHAP est capable de calculer les scores de contribution des variables indifféremment de la nature de ces modèles.

Le problème avec cette méthode dans le flux de données est que pour expliquer une donnée, elle a besoin de construire plusieurs autres données voisines de la donnée à expliquer afin d'approximer localement le modèle à expliquer ; ce qui peut ajouter un délai supplémentaire au moment de produire des explications. Le modèle qui sert à expliquer est également mis à jour après chaque taille de fenêtre au même titre que le modèle de détection d'anomalies. Les explications produites sont également sur un format similaire au format du flux et sont constituées de l'instant de mesure de la donnée et pour chaque variable de sa contribution. Cette considération permettra de visualiser les explications de manière assez intuitive.

3.4 Module de visualisation

Le flux de données, les contributions des variables ainsi que les scores d'anomalie sont présentés de manière intuitive afin de permettre à l'utilisateur de voir sur un même graphique, la courbe du flux, les anomalies ainsi que les contributions des variables en fonction du temps. Cette présentation permet à un utilisateur d'examiner facilement le comportement des variables

2. Le nombre de points (données) dans la fenêtre est égal à la taille de la fenêtre.

mises en avant par la méthode d'explication et d'observer en même temps le comportement du flux pour voir s'il y a matière à traiter l'alerte.

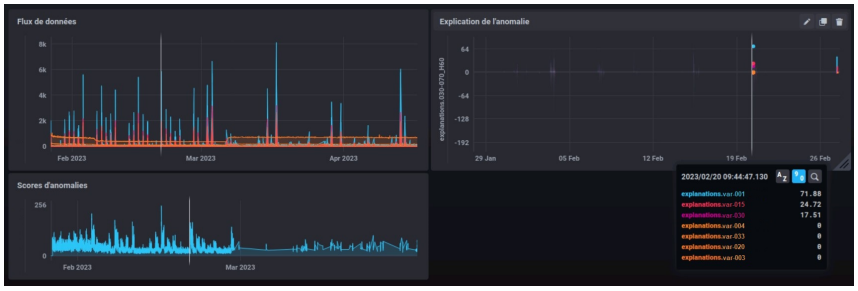


FIG. 2 – Visualisation comparée : flux, anomalies, explications.

La figure 2 présente une vue comparée du flux, des scores d'anomalie et des explications en fonction du temps. En observant les trois graphiques de la figure à un instant donné, on peut voir les variables qui ont le plus contribué au score d'anomalie et voir en même temps le comportement du flux à cette date.

4 Validation du système

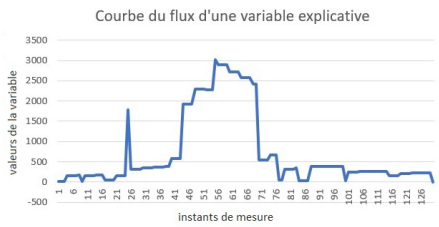
Le système mis en place a été utilisé pour détecter et expliquer les anomalies à partir d'un flux de données multivariées simulé à partir des données réelles contenues dans un fichier CSV. La validation a principalement consisté à observer à partir des explications fournies par la méthode d'explicabilité, les variables à plus forte contribution comme dans la figure 3c où on a les scores de la variable la plus contributive pendant une succession d'instants où le score d'anomalie était élevé (figure 3b). Ensuite, on peut aller au niveau du flux (figure 3a) pour observer le comportement de la variable à ces mêmes instants.

On peut constater à partir de ces 3 figures, une cohérence entre les scores d'anomalie, la tendance du flux et les scores de contribution de la variable ayant eu la plus forte contribution. Dans les figures 3a, 3b et 3c, les entiers au niveau de l'axe des abscisses représentent les instants de mesure, et les ordonnées respectivement les valeurs de mesure de la variable, les scores d'anomalie et les contributions à chaque instant.

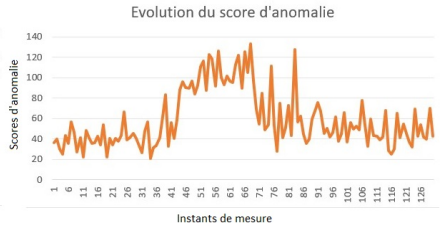
5 Conclusion et perspectives

Dans ce travail [3], nous avons présenté un prototype d'un système de détection et d'explication des anomalies en temps réel. Le sytème mis en place est composé de quatre modules essentiels à savoir : un module de détection des anomalies, une base de données de série temporelle, un module d'explicabilité et un outil de visualisation. Les résultats obtenus montrent comment l'explicabilité apporte une information importante aux experts du domaine pour mener des investigations en cas de détection d'une anomalie.

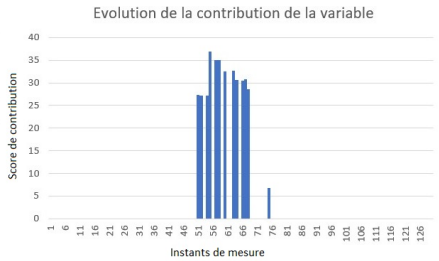
3. Nous remercions la Bpifrance, l'Isite CAP 20-25 et le MESR pour le financement de ce travail



(a) Flux de la variable la plus contributive.



(b) Evolution du score d'anomalie global.



(c) Score de la variable la plus contributive.

FIG. 3 – Vue comparée : valeur de la variable la plus contributive au score d'anomalie, score d'anomalie global, et contribution de la variable à chaque instant.

En guise de perspectives, nous souhaitons poursuivre la validation opérationnelle de ce prototype, aussi bien sur les techniques de détection des anomalies en continu, que sur les outils d'explicabilité pour une meilleure prise en compte de l'aspect temps réel.

Références

Arslan, F., A. Javaid, M. Danish Zaheer Awan, et Ebad-ur-Rehman (2023). Anomaly detection in time series : Current focus and future challenges. In D. V. K. Parimala (Ed.), *Anomaly Detection - Recent Advances, AI and ML Persp. and Appli.*, Chapter 3. Rijeka : IntechOpen.

Binder, A., S. Bach, G. Montavon, K. Müller, et W. Samek (2016). Layer-wise relevance propagation for deep neural network architectures. In K. J. Kim et N. Joukov (Eds.), *Information Science and Applications (ICISA) 2016*, Singapore, pp. 913–922. Springer Singapore.

Groen, A. M., R. Kraan, S. F. Amirkhan, J. G. Daams, et M. Maas (2022). A systematic review on the use of explainability in deep learning systems for computer aided diagnosis in radiology : Limited use of explainable ai ? *European Journal of Radiology 157*, 110592.

Khelifati, A., M. Khayati, A. Dignös, D. Difallah, et P. Cudré-Mauroux (2023). Tsm-bench : Benchmarking time series database systems for monitoring applications. *Proc. VLDB Endow. 16*(11), 3363–3376.

Kingma, D. P. et M. Welling (2019). An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning 12*(4), 307–392.

Lundberg, S. M. et S. Lee (2017). A unified approach to interpreting model predictions. In *Proc. of the 31st Intl. Conf. on Neural Information Processing Systems*, NIPS, Red Hook, NY, USA, pp. 4768–4777. Curran Associates Inc.

Mansalis, S., E. Ntoutsi, N. Pelekis, et Y. Theodoridis (2018). An evaluation of data stream clustering algorithms. *Statistical Analysis and Data Mining : The ASA Data Science Journal 11*(4), 167–187.

Mensi, A. et M. Bicego (2021). Enhanced anomaly scores for isolation forests. *Pattern Recognition 120*, 108115.

Mirsky, Y., T. Doitshman, Y. Elovici, et A. Shabtai (2018). Kitsune : An ensemble of autoencoders for online network intrusion detection. *ArXiv abs/1802.09089*.

Ngo Bibinbe, A. M. S., A. J. Djiberou Mahamadou, M. F. Mbouopda, et E. Mephu Nguifo (2022a). Dragstream : An anomaly and concept drift detector in univariate data streams. In *IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 842–851.

Ngo Bibinbe, A. M. S., M. F. Mbouopda, G. R. Mbiadou Saleu, et E. Mephu Nguifo (2022b). A survey on unsupervised learning algorithms for detecting abnormal points in streaming data. In *Intl. Joint Conf. on Neural Networks, IJCNN, Padua, Italy, July 18-23*, pp. 1–8.

Ribeiro, M. T., S. Singh, et C. Guestrin (2016). "why should i trust you ?" : Explaining the predictions of any classifier. In *Proc. of 22nd ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*, KDD, New York, NY, USA, pp. 1135–1144. ACM.

Tan, S. C., K. M. Ting, et T. F. Liu (2011). Fast anomaly detection for streaming data. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two*, IJCAI'11, pp. 1511–1516. AAAI Press.

Summary

Anomaly detection refers to the identification of rare events that differ significantly from the normal behavior observed in the data distribution. When the number of variables to analyze is large, it can be difficult to understand why the system has fired an alert without explanation. In this work, we present the prototype of an explainable real-time monitoring and anomaly detection system, on measurements obtained from a data stream. The built system consists of a set of anomaly detection methods combining deeplearning and decision trees as well as an agnostic explainability method. In an unsupervised learning context, we also show how explainability provides insights to validate the system along with feedback from domain experts.