



HAL
open science

Explicabilité de séries temporelles : étude de cas sur des données de sécurité urbaine

Matthieu Delahaye, Lina Fahed, Florent Castagnino, Philippe Lenca

► To cite this version:

Matthieu Delahaye, Lina Fahed, Florent Castagnino, Philippe Lenca. Explicabilité de séries temporelles : étude de cas sur des données de sécurité urbaine. Atelier EXPLAIN'AI @ Extraction et Gestion des Connaissances (EGC), Jan 2025, Strasbourg, France. hal-04955005

HAL Id: hal-04955005

<https://hal.science/hal-04955005v1>

Submitted on 18 Feb 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Explicabilité de séries temporelles : étude de cas sur des données de sécurité urbaine

Matthieu Delahaye*, Lina Fahed*
Florent Castagnino**, Philippe Lenca*

*IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238 Brest, France

**IMT Atlantique, LEMNA, F-44307 Nantes, France
prénom.nom@imt-atlantique.fr

Résumé. Les méthodes d’explicabilité en intelligence artificielle (XAI) sont devenues essentielles pour mieux comprendre les modèles complexes notamment dans les domaines sensibles tels que la sécurité urbaine. Dans cet article, nous étudions un jeu de données public gouvernemental sur les crimes et délits commis en France et mettons en application une méthode d’explicabilité à base de shapelets, des sous-séquences distinctives qui ressortent les informations pertinentes de séries temporelles d’infractions. Après avoir transformé le problème pour une tâche de classification supervisée, nous cherchons à faire apparaître des motifs caractéristiques de la criminalité. Néanmoins, la qualité et la disponibilité des données représentent un défi majeur pour la performance des modèles et leur interprétation. Ces conclusions illustrent le besoin d’ouvrir davantage les ressources et les données afin de mieux comprendre et analyser les phénomènes de sécurité urbaine, un milieu encore trop peu accessible.

1 Introduction

Les modèles d’apprentissage automatique, en particulier ceux dits “boîtes noires”, se sont imposés dans divers domaines grâce à leurs résultats remarquables. Cependant, leur complexité croissante, motivée par la recherche de performances, a entraîné un manque de transparence, limitant ainsi la compréhension de leurs utilisateurs. Face à cette problématique, le domaine de l’Intelligence Artificielle Explicable (XAI) a émergé, donnant naissance à diverses méthodes d’interprétation et d’explication des résultats des modèles (Adadi et Berrada, 2018).

L’explicabilité est un enjeu dans des domaines critiques tels que la défense, la médecine ou la sécurité, où les décisions ont des conséquences majeures. Dans ce contexte, les utilisateurs/contributeurs peuvent souhaiter appréhender les modèles d’apprentissage automatique d’un bout à l’autre, à la recherche d’une cohérence entre leurs entrées et sorties (Barredo Arrieta et al., 2020). Ce besoin de transparence est aussi influencé par le type de données traitées. Si la recherche en explicabilité s’est surtout orientée sur des données intuitives comme les images et le texte, les séries temporelles représentent un défi supplémentaire : leur interprétation n’est pas évidente, un signal qui évolue dans le temps nous est bien moins familier que des images ou des mots (Rojat et al., 2021). Pourtant, les séries temporelles, qui représentent l’objet de notre étude, sont omniprésentes dès lors que la notion de temps entre en jeu.

Dans la littérature, nous distinguons plusieurs familles de méthodes d’XAI, notamment, les méthodes *post-hoc* qui expliquent les résultats ou la structure des modèles d’apprentissage, et les méthodes *ante-hoc* qui sont transparentes par conception (Guidotti et al., 2019). Une distinction est faite sur la nature du modèle d’apprentissage : les méthodes *agnostiques* sont adaptées à n’importe quel type de modèles indépendamment de son architecture, tandis que d’autres sont *spécifiques* à une famille de modèles. Par ailleurs, certaines méthodes d’XAI sont développées pour un type de données particulier alors que d’autres sont plus génériques.

Nous nous focalisons sur les méthodes d’XAI appliquées aux séries temporelles. L’approche des shapelets, introduite par Ye et Keogh (2009), se distingue par sa capacité à extraire des sous-séquences représentatives et discriminantes des séries temporelles. Elle offre la possibilité de fournir des explications sous forme de motifs directement issus des données brutes. L’algorithme *Learning Shapelets* de Grabocka et al. (2014) améliore cette approche en intégrant un cadre d’apprentissage qui évite de devoir parcourir l’ensemble des sous-séquences au sein des séries, tout en combinant précision et interprétabilité.

Nous appliquons l’algorithme *Learning Shapelets* à un domaine à fort impact social et particulièrement sensible, à savoir la sécurité urbaine, qui nécessite des modèles à la fois performants et interprétables. **La contribution de cet article réside dans le traitement étendu de données de criminalité et de délinquance et l’identification des motifs temporels, de type shapelets, les plus représentatifs afin d’obtenir de nouvelles clés de compréhension du domaine complexe de la sécurité urbaine.**

Dans la section 2, nous passons en revue les travaux principaux de l’état de l’art sur les approches d’XAI pour les séries temporelles. Nous introduisons notre cas d’étude dans la section 3, puis, dans la section 4, nous présentons l’application du modèle d’XAI retenu. Nous discutons des résultats de notre étude dans la section 5. Enfin, dans la section 6, nous résumons les apports de l’étude et présentons les perspectives de recherche futures.

2 État de l’art

Nous présentons ci-dessous une étude de travaux en XAI portant sur les séries temporelles ainsi que des définitions inspirées de l’état de l’art.

Selon Rojat et al. (2021), l’explication devient nécessaire lorsque la tâche à accomplir dépasse les capacités d’un modèle simple et interprétable, tout en étant trop importante pour être confiée à un modèle opaque dont les décisions seraient difficiles à justifier et à accepter. Pour répondre à ce besoin, l’XAI cherche à fournir des explications des modèles et/ou des résultats afin de permettre aux utilisateurs de se les approprier au mieux.

L’explicabilité des séries temporelles pose des défis importants en raison de leur nature séquentielle et non intuitive, ainsi que par des relations temporelles complexes qui existent entre chaque point de mesure de la série (Rojat et al., 2021). Certaines méthodes, utilisées principalement sur des données intuitives (*e.g.* texte, images), ont pu être réadaptées pour des séries temporelles (Ge et al., 2018; Vinayavekhin et al., 2018; Kashiparekh et al., 2019). Par ailleurs, d’autres approches se sont développées pour répondre directement aux défis des séries temporelles notamment en identifiant des motifs et des tendances au sein des séries (Lin et al., 2007; Ye et Keogh, 2009). Nous explorons diverses méthodes d’XAI appliquées aux séries temporelles (voir définition 2.1), en débutant par celles développées sur d’autres types de données, puis par celles conçues pour les séries temporelles.

Définition 2.1 Une **série temporelle** $T = \{x_1, x_2, \dots, x_M\} \in \mathbb{R}^{M \times D}$ est un ensemble ordonné de M observations réelles de dimension D . Un jeu de données $J = \{T_1, T_2, \dots, T_N\}$ est une collection de N séries temporelles où $J \in \mathbb{R}^{N \times M \times D}$. Lorsque $D = 1$ les séries temporelles sont dites univariées tandis que lorsque $D > 1$ les séries sont dites multivariées.

À partir de modèles de classification tels que les *Convolutional Neural Network (CNN)*, les *Recurrent Neural Network (RNN)* ou les *Transformers*, diverses méthodes d’XAI se sont développées. Pour les CNN, citons la méthode de perturbation *ConvTimeNet* de Kashiparekh et al. (2019) qui se base sur l’*“occlusion sensitivity”*. En perturbant les données par occlusion, il est possible de retrouver le niveau de contribution des parties de l’image à sa classification par le modèle. Des modèles de traitement de texte tels que les *Long Short Term Memory (LSTM)* sont particulièrement efficaces en raison de leur capacité à traiter les données de manière séquentielle. Par exemple, l’étude de Ge et al. (2018) combine un CNN avec un LSTM, puis, un réseau de neurones feedforward est utilisé pour la classification, dont les poids de sa dernière couche indiquent l’importance de chaque observation de la série temporelle. Des mécanismes d’attention, utilisés dans les *transformers* notamment, sont aussi proposés (Vinayavekhin et al., 2018) afin de concentrer le modèle sur les parties importantes de la série temporelle. Cependant, différentes attentions peuvent aboutir au même résultat risquant de donner des explications trompeuses (Jain et Wallace, 2019). Des techniques comme *Local Interpretable Model-agnostic Explanation (LIME)* (Ribeiro et al., 2016) et *SHapley Additive Explanations (SHAP)* (Lundberg et Lee, 2017) peuvent être adaptées pour traiter les séries temporelles selon chaque observation réelle (voir définition 2.1). Cependant, prendre en compte chaque point de la série temporelle comme une variable présente un défi : l’hypothèse de l’indépendance des variables n’est pas respectée en raison des observations temporellement adjacentes de la série (Watson, 2022). Ces méthodes présentent donc plusieurs limites pour les séries temporelles compliquant l’interprétation de leurs résultats.

D’autres approches comme l’explicabilité par l’exemple (Ming et al., 2019) ou par contre-factuels (Delaney et al., 2021), ont été développées, mais présentent des contraintes : (i) elles sont basées sur l’intuition humaine ce qui est parfois mis en difficulté par la nature même des séries temporelles et (ii) la capacité à fournir un exemple interprétable par l’utilisateur dépend fortement de son niveau de compréhension (Theissler et al., 2022).

Les méthodes *features-based* (Ito et Chakraborty, 2019) permettent d’extraire des caractéristiques à partir des séries temporelles brutes de manière à en déduire un sens ou améliorer la précision du modèle. Néanmoins, ces méthodes ne sont pas forcément adaptées aux utilisateurs cibles et dépendent aussi de la capacité d’interprétabilité des caractéristiques.

En comparaison, des travaux sur des méthodes d’XAI conçues spécialement pour identifier les éléments distinctifs des séries temporelles suscitent beaucoup d’intérêt. Nous allons présenter deux exemples de ces méthodes.

Méthode SAX : La méthode *Symbolic Aggregate approXimation (SAX)* (Lin et al., 2007) discrétise une série temporelle en une suite de lettres. Des symboles (*i.e.* lettres) issus de la discrétisation sont assignés à chacun des segments et forment des mots. Ces mots représentent des motifs dont nous pouvons analyser la fréquence afin d’en tirer une interprétation.

Shapelets : Les méthodes d’extraction de shapelets introduites par Ye et Keogh (2009) ainsi que celle de Grabocka et al. (2014) ont retenu notre attention.

Définition 2.2 Un **shapelet** de taille L est une séquence ordonnée de valeurs. Sémantiquement, le shapelet a une valeur informative importante car il doit permettre de séparer facilement les classes au sein du jeu de données. Les K shapelets les plus informatifs sont notés $S \in \mathbb{R}^{K \times L}$.

Dans la méthode initiale de Ye et Keogh (2009), des sous-séquences des séries temporelles sont d’abord récupérées, et définies comme possible *candidat* à appartenir à S puis, la distance entre chaque *candidat* et toutes les sous-séquences de même taille dans chaque série est mesurée. La distance minimale, caractéristique de la meilleure correspondance entre une série et un shapelet, devient la distance finale entre l’intégralité de chacune des séries temporelles et chacun des candidats. Les shapelets finaux sont les *candidats* qui maximisent le gain d’information de séparation des classes (voir définition 2.2). L’algorithme *Learning Shapelets* de Grabocka et al. (2014) intègre une fonction d’apprentissage pour définir les shapelets, évitant ainsi le traitement l’ensemble des sous-séquences ce qui réduit significativement les temps de calcul. De par ses qualités de performance et d’interprétabilité, la méthode *Learning Shapelets* est, pour nous, la plus appropriée pour notre cas d’étude.

3 Cas d’étude : sécurité urbaine

Notre étude dans le domaine de la sécurité urbaine s’inscrit dans le cadre d’un projet ANR¹ dont un objectif est d’aider les services de police et de gendarmerie à accepter davantage les outils d’intelligence artificielle en leur apportant une compréhension des phénomènes de criminalité et de délinquance à l’aide de modèles d’apprentissage automatique explicable. Ce travail est mené avec des experts en sciences sociales travaillant sur les phénomènes de criminalité.

En 1972, le gouvernement français a mis en place l’“État 4001”, un outil de mesures statistiques mensuelles des crimes et des délits². Ces mesures correspondent aux plaintes et signalements enregistrés par les services de police et de gendarmerie nationales à travers les départements français selon 107 catégories d’infractions entre 1996 et 2022, distribuées via un jeu de données public. Dans la suite de cette section, nous décrivons ce jeu de données ainsi que son adaptation à une tâche de classification supervisée.

3.1 Description du jeu de données

Dans le jeu de données de l’État 4001³, les 107 catégories de crimes et de délits sont transformées en 10 variables : *Homicides*, *Vols*, *Cambriolages/violation de domicile*, *Crimes/délits sexuels*, *Infractions liées aux stupéfiants*, *Dégradations/destructions de biens volontaires*, *Violences*, *Falsifications et contrefaçons*, *Délits Économiques et Financiers* et *Irrégularités de Main-d’œuvre*. Les données sont séquencées pour chaque année en 12 mois et par département selon les 10 variables et forment un ensemble de 2193 séries temporelles dans un jeu de données $J \in \mathbb{R}^{2193 \times 12 \times 10}$ (voir définition 2.1).

1. <https://anr.fr/Projet-ANR-21-CE26-0023>

2. <https://mobile.interieur.gouv.fr/Interstats/Sources-et-methodes-statistiques/Glossaire/Etat-4001>

3. <https://www.data.gouv.fr/fr/datasets/chiffres-departementaux-mensuels-relatifs-aux-crimes-et-delits-enregistres-par-les-services-de-police-et-de-gendarmerie-depuis-janvier-1996/>

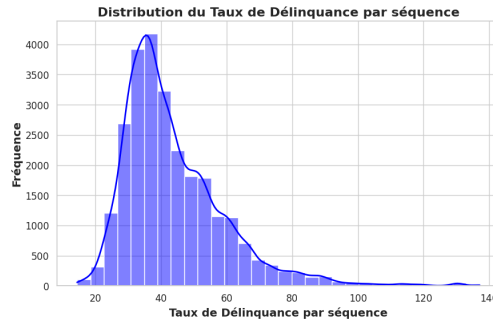


FIG. 1 – Distribution du taux de délinquance pour l'ensemble des séries temporelles.

3.2 Adaptation pour une tâche de classification

Après concertation avec les experts en sciences sociales travaillant sur la criminalité, nous souhaitons obtenir une prédiction du niveau de délinquance pour chaque département. L'adaptation du jeu de données (création d'une variable cible) pour une tâche de classification supervisée permet de répondre à l'objectif d'explicabilité par l'identification de motifs temporels distinctifs liés avec le niveau de criminalité. Nous mesurons le taux de délinquance de chacune des séries temporelles, qui correspond au rapport entre la somme des valeurs de chaque catégorie de délits/crimes et de la population du département sur l'année de chaque série, le tout multiplié par un facteur 1000. Les séries temporelles représentent des séquences annuelles, nous prenons la somme des catégories des 12 observations mensuelles de chaque série afin d'avoir un résultat à l'année.

La distribution du taux de délinquance des séries temporelles (voir figure 1), est asymétrique avec une longue queue vers la droite qui témoigne d'un grand nombre de valeurs élevées éloignées de la moyenne. L'hypothèse de normalité est rejetée par les tests de Shapiro-Wilk et de Kolmogorov-Smirnov.

Afin de définir les classes pour labelliser nos séries temporelles, nous avons étudié l'adaptabilité de deux méthodes de discrétisation du taux de délinquance et une méthode ensembliste :

- *Répartition par intervalles égaux* : n intervalles de même taille sont définis par rapport à l'étendue des valeurs de notre variable cible *i.e.* taux de délinquance, puis, une étiquette est assignée en fonction de chaque intervalle. Cependant, la distribution du taux de délinquance n'étant pas symétrique, la répartition par intervalles égaux risque de former des classes déséquilibrées et biaiser les résultats de la classification.
- *Clustering par K-Means* : assigne une étiquette pour un nombre k de clusters. Néanmoins, les clusters sont définis à partir de l'ensemble des variables d'intérêts, donc sur un espace multidimensionnel, ce qui complexifie leur interprétation.
- *Répartition par quantiles* : le jeu de données est découpé en n parties proportionnelles, avec le même nombre de séries temporelles par partie, chaque partie se voit assigner un label. Cette répartition s'adapte à la distribution du taux de délinquance afin de produire des classes équilibrées.

La différence de proportion de séries temporelles selon chaque méthode de labellisation est mise en évidence figure 2 pour une répartition en 4 classes. Les méthodes de répartition par

Explicabilité de séries temporelles pour la sécurité urbaine

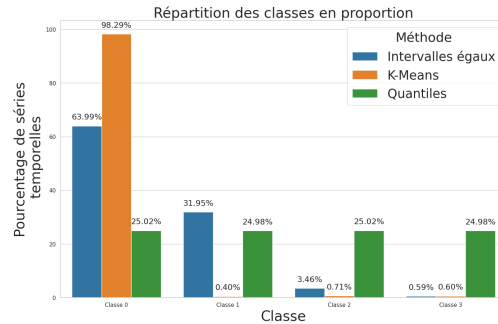


FIG. 2 – Comparaison de la proportion de séries temporelles de chaque classe selon trois méthodes de labellisation.

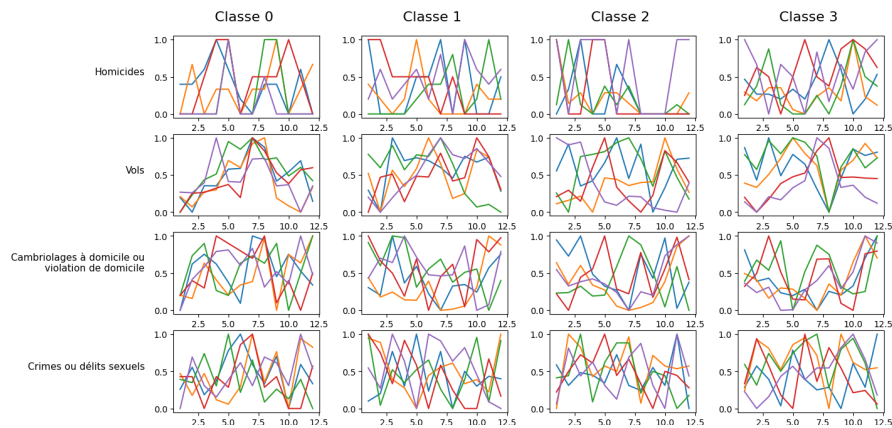


FIG. 3 – Représentation des séries temporelles propres à 4 classes selon différentes variables.

intervalles égaux et d'utilisation de K-Means entraînent un déséquilibre important des classes comparé à la répartition par quantiles dont la répartition est similaire pour chaque classe. De plus, avec la répartition par quantiles, chaque classe est définie selon des seuils fixes (*i.e.* valeur min et max par classe) aisément compréhensibles par l'utilisateur. Grâce à son adaptation à toute distribution et sa facilité d'explication, la répartition par quantiles est l'approche la plus adaptée à notre cas d'étude. Dans la suite de l'étude, les séries temporelles seront réparties selon 4 classes, ce qui revient à diviser en quartiles. Cela permet une analyse comparative entre les classes facilitant l'interprétation.

Le jeu de données est découpé en un jeu d'entraînement (80%), validation (10%) et de test (10%). Comme nous cherchons à comparer les formes relatives des séries temporelles, les séries sont standardisées indépendamment sur une moyenne nulle et un écart-type unitaire en premier lieu sur le jeu d'entraînement puis sur le jeu de test, puis normalisées sur la plage [0,1] afin de simplifier les calculs de distances et l'interprétation globale.

Dans la figure 3, pour chaque classe, les séries sont représentées selon différentes variables. Ainsi, nous remarquons que la série temporelle rouge de la classe 2 a un pic marqué sur chaque

variable au mois de mai, significatif d'une hausse de criminalité. Néanmoins, aucune tendance ne se distingue sur les variables, c'est-à-dire que nous ne retrouvons pas de motif spécifique à une catégorie. Cela peut être dû à la taille réduite des séries temporelles étudiées (*i.e.* 12 observations). La répartition des classes étant faite par quartiles, nous pouvons interpréter graduellement chaque classe par niveau de délinquance : la classe 0 étant caractéristique du niveau de délinquance le plus faible. Maintenant que les données sont labellisées et pré-traitées, nous pouvons appliquer l'algorithme retenu *Learning Shapelets*.

4 Déploiement du modèle

Nous présentons ici le déploiement de l'algorithme *Learning Shapelets* puis l'entraînement d'un modèle de classification supervisée sur les données transformées par l'algorithme.

Learning Shapelets : *Learning Shapelets* de Grabocka et al. (2014) est un algorithme d'apprentissage basé sur la méthode des shapelets de Ye et Keogh (2009). Les shapelets sont déterminés (voir définition 2.2) à partir d'une fonction objectif. Nous utilisons le package Python *tslearn* (Tavenard et al., 2020) pour déployer l'algorithme de *Learning Shapelets*.

Définition 4.1 Distance entre un shapelet et une série temporelle : La distance entre la n -ième série temporelle T_n et le k -ième shapelet S_k correspond à la distance minimale $M_{i,k}$ entre S_k et les différents segments i extraits de T_n , afin de calculer la ressemblance entre un shapelet et le segment de la série qui est le plus proche de ce dernier (Grabocka et al., 2014).

Définition 4.2 Transformation de l'espace de données : Les distances minimales aux shapelets permettent de transformer les séries temporelles $T \in \mathbb{R}^{N \times M}$ en une nouvelle représentation $M \in \mathbb{R}^{N \times K}$ (Lines et al., 2012; Grabocka et al., 2014)

Fonctionnement de *Learning Shapelets* : Tout d'abord, un algorithme K-Means est appliqué à l'ensemble des segments des séries temporelles brutes. Les centroïdes représentent les shapelets initiaux. Un nombre de shapelets par taille est déterminé par l'algorithme *Learning Shapelets*, en fonction du nombre de séries temporelles et de classes.

Pour chaque shapelet, la distance euclidienne minimale à tous les segments de chaque série temporelle est mesurée (voir définition 4.1). La fonction *min* n'étant pas différentiable, elle est remplacée par une fonction différentiable approximée *soft-min*, afin d'utiliser un algorithme d'optimisation pour le modèle supervisé. Par la suite, les séries temporelles sont transformées comme présenté définition 4.2, puis, une régression logistique est appliquée sur le nouvel espace afin de classer les séries. La fonction de perte associée est une *log-loss* régularisée. Un algorithme d'optimisation met à jour les valeurs des shapelets et des poids du modèle linéaire. Pour la suite de l'étude, nous avons choisi l'algorithme Adam (Kingma, 2014), plus performant que la descente de gradient stochastique.

Récupération des attributs : L'algorithme est déployé pour chaque variable indépendamment, ainsi, les shapelets appris correspondent à une seule variable. Les espaces transformés propres à chaque variable sont concaténés de manière à former un seul et même jeu de données transformé pour l'ensemble des variables. En complément, pour chaque variable, nous

Explicabilité de séries temporelles pour la sécurité urbaine

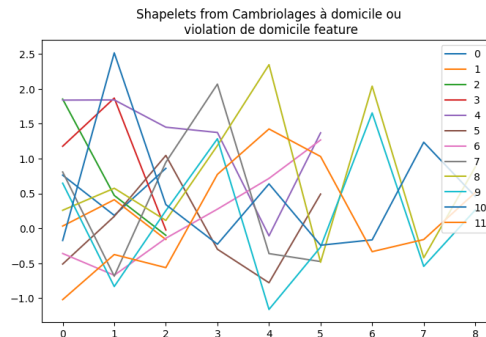


FIG. 4 – Ensemble des 12 shapelets de la variable “Cambriolages à domicile ou violation de domicile” résultant de l’algorithme *Learning Shapelets* : 4 shapelets de taille 3, 4 shapelets de taille 6 et 4 shapelets de taille 9.

récupérons la forme de chaque shapelet, les paramètres de la dernière couche d’apprentissage du modèle responsable de la classification et la localisation de la sous-séquence dont la distance est minimale avec le shapelet. Ces attributs vont nous aider à interpréter les résultats par la suite. L’ensemble des 12 shapelets générés sur la variable “Cambriolages à domicile ou violation de domicile” sont visibles sur la figure 4.

Classification à partir de l’espace transformé : Nous constatons que *Learning Shapelets* n’arrive pas à séparer nettement les classes entre elles, les séries temporelles transformées forment un ensemble de points qui se superposent sans groupes distincts (voir figure 5). Pour cette raison, et après avoir étudié plusieurs modèles de classification, nous optons pour le modèle *Random Forest* (Breiman, 2001).

5 Résultats et discussions

Dans cette section, nous décrivons les shapelets obtenus, les résultats de classification du modèle ainsi que les interprétations qui en découlent.

Limites de décision : *Learning Shapelets* génère des shapelets de façon à ce que : (i) leurs distances avec les séries temporelles soient séparées linéairement en fonction de leur classe, (ii) les distances avec les séries temporelles de chaque classe soient regroupées ensemble et éloignées de celles des séries des autres classes. Pour chaque variable, nous pouvons visualiser ces concepts en utilisant les paramètres de la couche de classification de *Learning Shapelets*.

L’analyse de l’importance de caractéristiques du *Random Forest* permet d’identifier les shapelets les plus influents pour une variable. La figure 5 présente les deux shapelets les plus significatifs : shapelets 5 et 6 (représentés en haut à gauche de la figure) de la variable “Vols”. Ils sont de même taille mais capturent une tendance différente. Les séries temporelles de chaque classe sont représentées sur des graphiques au milieu et en bas gauche de la figure 5. Bien

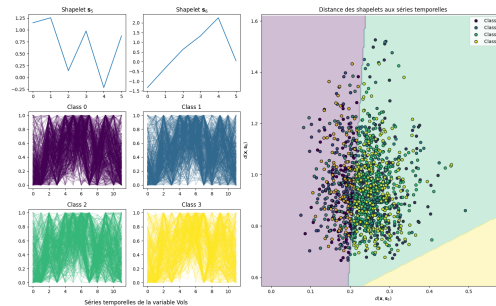


FIG. 5 – Distances entre les séries temporelles et les shapelets 5 (*abscisses*) et 6 (*ordonnées*) correspondant à la variable “Vols”.

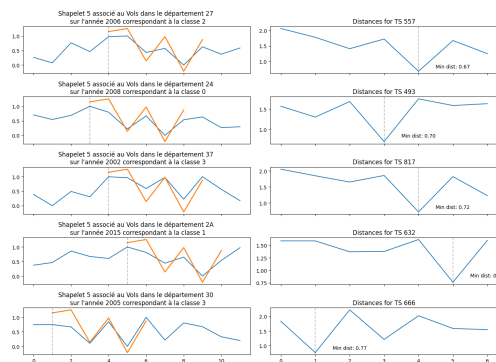


FIG. 6 – Localisation des séries temporelles dont les distances sont les plus proches avec le cinquième shapelet de la variable “Vols”.

qu’une tendance d’observations élevées entre les mois de mai et de juillet se démarque pour l’ensemble des séries, nous ne distinguons pas de forme particulière au sein des classes.

Le graphique à droite de la figure 5 permet de visualiser les séries temporelles transformées selon leurs distances aux deux shapelets (*i.e.* shapelet 5 en abscisses et 6 en ordonnées). À l’aide des paramètres de la couche de classification, nous pouvons tracer les limites de décision linéaires visibles sur le graphique à droite de la figure 5. Ces paramètres rendent compte de l’importance des distances de chaque shapelet avec l’ensemble des séries. Ils reflètent la contribution de chaque shapelet à la séparation des classes dans l’espace transformé ce qui donne une piste d’interprétation sur l’influence des shapelets identifiés. Cependant, les points de la classe 0 se retrouvent à la fois dans les limites de la classe 0 et celles de la classe 2. Le phénomène est similaire pour les points de la classe 2. De plus, aucun point de la classe 3 ne se retrouve dans sa limite de décision. Les données se regroupent en un ensemble de points, avec des séries de chaque classe qui se superposent, ce qui complique la séparation linéaire.

Localiser les correspondances des shapelets avec les séries temporelles : La figure 6 illustre le shapelet 5 associé à la variable “Vols”, le shapelet dont le score d’importance est

le plus élevé dans le *Random Forest*. Ce shapelet, de taille 6, traduit une tendance sur une période 6 mois. Les graphiques à gauche de la figure 6 représentent les 5 séries temporelles (en bleu) dont les distances sont les plus courtes avec le shapelet étudié (en orange). La forme du shapelet est globalement similaire à celle des segments des séries temporelles représentées. Les graphiques à droite correspondent à la distance entre le shapelet et chaque segment de la série temporelle étudiée. Par exemple, pour la première série en haut du graphique, la distance est minimale au cinquième point, (mesuré sur le cinquième segment de la série temporelle). Ce point est marqué par une ligne grise pointillée sur les deux graphes de chaque série.

Par ailleurs, pour les quatre premières séries temporelles, la distance est minimale sur des segments proches : 3, 4 et 5, couvrant la période de avril à novembre. Nous cherchons à savoir si cette période est significative et caractérisée par d'autres informations : année, département et classe de la série temporelle. Les résultats sont peu concluants : les années et les départements sont dispersés. Nous manquons d'informations pour déterminer la période ou le lieu représentatif de ce shapelet. Plus étonnamment, aucune classe ne se distingue alors que l'objectif de *Learning Shapelets* est justement de déterminer des shapelets caractéristiques d'une classe. Nous pouvons conclure que les shapelets appris par l'algorithme ne ressortent pas d'informations distinctives des séries temporelles, ce qui complique l'interprétation des motifs obtenus. Le modèle semble être très indécis sur le choix de la classe (accuracy = 0.300). La matrice de confusion montre qu'il classifie presque uniformément les séries sur chacune des classes. Les shapelets générés n'ont donc pas séparé les classes efficacement.

6 Discussion et conclusion

Nous avons étudié un domaine à fort impact social, à savoir la sécurité urbaine, dans l'objectif de mieux comprendre les phénomènes de criminalité et leurs tendances en appliquant un algorithme d'XAI à base de shapelets sur des données publiques. Nous avons adapté les données à une tâche de classification supervisée en introduisant une variable cible afin de rendre l'interprétation des shapelets plus accessible, notamment aux sociologues. Nous soulignons l'importance de rendre les données de sécurité urbaine accessibles et complètes, tout en favorisant le développement de méthodes d'explicabilité adaptées à d'autres contextes critiques.

Le principal défi de notre étude réside dans la quantité et la qualité des données disponibles. Le jeu de données provenant de l'“État 4001” représente uniquement les plaintes déposées, négligeant la délinquance non signalée. De plus, malgré la publication de données publiques, les ressources fournies sont limitées car très agrégées ce qui complique leur interprétation. L'autre limite de notre étude est la segmentation des séries temporelles par année qui fait perdre la dépendance temporelle entre les séries et empêche l'étude des tendances à long terme.

Pour répondre à ces défis, il est possible d'utiliser des jeux de données plus élaborés pour fiabiliser les résultats tout en conservant une analyse des tendances sur la durée, telle que la base de données couplant les mesures de l'“État 4001” avec des enquêtes de victimation⁴ réalisées auprès des ménages. Cependant, les mesures sont disponibles par année uniquement ce qui contraint l'analyse précise des phénomènes de criminalité. Une solution est de s'intéresser aux pays dont la politique d'accessibilité est plus ouverte. La police de Toronto⁵, par exemple,

4. Enquête Vécu et Ressenti en matière de Sécurité (VRS) : <https://mobile.interieur.gouv.fr/Interstats/L-enquete-Vecu-et-ressenti-en-matiere-de-securite-VRS>

5. Données sur les infractions à Toronto : <https://data.torontopolice.on.ca/search?collection=Dataset&q=crime>

publie des données horaires sur la commission des infractions au sein de la ville, permettant des analyses à granularité temporelle variable.

Nous envisageons le déploiement de la méthode de classification non-supervisée par *unsupervised shapelet* (Zakaria et al., 2012) qui permet de s’affranchir de la dépendance à une classe. Afin d’apporter une compréhension plus globale au modèle, la détermination de shapelets sur des séries multivariées (Medico et al., 2021) ou bien l’intégration de connaissances au modèle (Wan et al., 2024) sont des perspectives prometteuses.

Références

- Adadi, A. et M. Berrada (2018). Peeking inside the black-box : A survey on explainable artificial intelligence (XAI). *IEEE Access* 6, 52138–52160.
- Barredo Arrieta, A., N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, et F. Herrera (2020). Explainable artificial intelligence (XAI) : Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58, 82–115.
- Breiman, L. (2001). Random forests. *Machine Learning* 45, 5–32.
- Delaney, E., D. Greene, et M. T. Keane (2021). Instance-based counterfactual explanations for time series classification. In *International conference on case-based reasoning*, pp. 32–47.
- Ge, W., J.-W. Huh, Y. R. Park, J. H. Lee, Y. Kim, et A. Turchin (2018). An interpretable ICU mortality prediction model based on logistic regression and recurrent neural networks with LSTM units. *AMIA Symposium* 2018, 460–469.
- Grabocka, J., N. Schilling, M. Wistuba, et L. Schmidt-Thieme (2014). Learning time-series shapelets. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 392–401.
- Guidotti, R., A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, et D. Pedreschi (2019). A survey of methods for explaining black box models. *ACM Computing Surveys* 51(5), 1–42.
- Ito, H. et B. Chakraborty (2019). A proposal for shape aware feature extraction for time series classification. In *2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST)*, pp. 1–6.
- Jain, S. et B. C. Wallace (2019). Attention is not Explanation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies, Volume 1*, pp. 3543–3556.
- Kashiparekh, K., J. Narwariya, P. Malhotra, L. Vig, et G. Shroff (2019). ConvtimeNet : A pre-trained deep convolutional neural network for time series classification. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8.
- Kingma, D. P. (2014). Adam : A method for stochastic optimization. *arXiv preprint arXiv :1412.6980*.
- Lin, J., E. Keogh, L. Wei, et S. Lonardi (2007). Experiencing SAX : A novel symbolic representation of time series. *Data Min. Knowl. Discov.* 15, 107–144.
- Lines, J., L. M. Davis, J. Hills, et A. Bagnall (2012). A shapelet transform for time series classification. In *Proceedings of the 18th ACM SIGKDD*, pp. 289–297.

- Lundberg, S. M. et S.-I. Lee (2017). A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 4768–4777.
- Medico, R., J. Ruysinck, D. Deschrijver, et T. Dhaene (2021). Learning multivariate shapelets with multi-layer neural networks for interpretable time-series classification. Advances in Data Analysis and Classification 15(4), 911–936.
- Ming, Y., P. Xu, H. Qu, et L. Ren (2019). Interpretable and steerable sequence learning via prototypes. In Proceedings of the 25th ACM SIGKDD, pp. 903–913.
- Ribeiro, M. T., S. Singh, et C. Guestrin (2016). "Why Should I Trust You?" : Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD, pp. 1135–1144.
- Rojat, T., R. Puget, D. Filliat, J. Del Ser, R. Gelin, et N. Díaz-Rodríguez (2021). Explainable artificial intelligence (XAI) on TimeSeries data : A survey. arXiv preprint arXiv :2104.00950.
- Tavenard, R., J. Faouzi, G. Vandewiele, F. Divo, G. Androz, C. Holtz, M. Payne, R. Yurchak, M. Rußwurm, K. Kolar, et E. Woods (2020). Tslern, a machine learning toolkit for time series data. Journal of Machine Learning Research 21(118), 1–6.
- Theissler, A., F. Spinnato, U. Schlegel, et R. Guidotti (2022). Explainable AI for time series classification : A review, taxonomy and research directions. IEEE Access 10, 100700–100724.
- Vinayavekhin, P., S. Chaudhury, A. Munawar, D. J. Agravante, G. De Magistris, D. Kimura, et R. Tachibana (2018). Focusing on what is relevant : Time-series learning and understanding using attention. In 2018 24th Int. Conf. on Pattern Recognition (ICPR), pp. 2624–2629.
- Wan, X., L. Cen, X. Chen, Y. Xie, et W. Gui (2024). Prior knowledge-augmented unsupervised shapelet learning for unknown abnormal working condition discovery in industrial process. Advanced Engineering Informatics 60, 102429.
- Watson, D. S. (2022). Conceptual challenges for interpretable machine learning. Synthese 200(2), 65.
- Ye, L. et E. Keogh (2009). Time series shapelets : a new primitive for data mining. In Proceedings of the 15th ACM SIGKDD, pp. 947–956.
- Zakaria, J., A. Mueen, et E. Keogh (2012). Clustering time series using unsupervised-shapelets. In 2012 IEEE 12th International Conference on Data Mining, pp. 785–794.

Summary

Explainability methods in artificial intelligence (XAI) have become essential for better understanding complex patterns, particularly in sensitive areas such as urban security. In this article, we study a governmental public dataset on crimes committed in France and implement an explainability method based on shapelets, distinctive subsequences that extract relevant information from time series of offenses. After transforming the problem for a supervised classification task, we aim to reveal patterns characteristic of crime. Nevertheless, data quality and availability represent a major challenge for model performance and interpretation. These findings illustrate the need to further open up resources and data in order to better understand and analyze urban security phenomena, an area that is not sufficiently accessible.