



HAL
open science

Apprentissage interprétable de la criminalité en France (2012-2021)

Nida Meddouri, David Beserra

► **To cite this version:**

Nida Meddouri, David Beserra. Apprentissage interprétable de la criminalité en France (2012-2021). Atelier GAST (Gestion et Analyse de données Spatiales et Temporelles), Nida Meddouri; Aurélie Leborgne; Loïc SALMON; Clément Iphar, Jan 2024, Dijon, France. hal-04875256

HAL Id: hal-04875256

<https://hal.science/hal-04875256v1>

Submitted on 8 Jan 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.

Apprentissage interprétable de la criminalité en France (2012-2021)

Nida Meddouri*, David Beserra*

* Laboratoire de Recherche de l'EPITA (LRE),
Le Kremlin-Bicêtre, France.
firstname.lastname@epita.fr,

L'activité criminelle en France a connu une évolution significative au cours des deux dernières décennies, marquée par la recrudescence des actes de malveillance, notamment liés aux mouvements sociaux et syndicaux, aux émeutes, ainsi qu'au terrorisme. Dans ce contexte difficile, l'utilisation de techniques issues de l'intelligence artificielle pourrait offrir de nombreuses perspectives pour renforcer la sûreté publique et privée en France. Un exemple de cette approche est l'analyse spatio-temporelle des données de criminalité, déjà couronnée de succès au Brésil (Da Silva et al., 2020), au Proche-Orient (Tolan et al., 2015), et dans d'autres pays. Dans le cadre de ce travail, nous explorons la possibilité d'appliquer cette approche au contexte français.

1 Données de criminalité en France (2012-2021)

Les données de criminalité en France, disponibles en ligne depuis le 9 octobre 2015 sous une licence ouverte (Open Licence)¹, offrent une vue complète des statistiques de crimes et délits enregistrés par les services nationaux de police et de gendarmerie. Ces données couvrent la période de 2012 à 2021 et englobent la France métropolitaine, les *Départements et Régions d'Outre-Mer (DROM)* ainsi que les *Collectivités d'Outre-Mer (COM)*.

Les informations sont déclinées en fonction des différents services de police et de gendarmerie qui les ont enregistrées, et elles proviennent de l'outil historique de mesure de l'activité des services, connu sous le nom "Etat 4001", en place depuis 1972. Ce dispositif statistique utilise une nomenclature comprenant 107 catégories d'infractions, avec 107 index au total, dont 4 ne sont pas utilisés. Cette base de données constitue ainsi une ressource essentielle pour comprendre et analyser l'évolution de la criminalité en France au fil du temps, et nous considérons que ces données peuvent être utilisées dans des analyses spatio-temporelles afin d'aider à "prévoir" et "interpréter" l'occurrence de certains types de crimes en France.

2 Analyse et pré-traitement des données criminelles en France

Avant d'utiliser ces données, il est nécessaire de les prétraiter afin de regrouper les statistiques fournies par les services de la police nationale avec celles de la gendarmerie nationale

1. <https://www.data.gouv.fr/fr/datasets/crimes-et-delits-enregistres-par-les-services-de-gendarmerie-et-de-police-depuis-2012/>

par année. Ensuite, étant donné que ces statistiques proviennent de 372 *Compagnies de Gendarmerie Départementale (CGD)* et 828 *Circonscriptions de Sécurité Publique (CSP)*, elles ont été regroupées par département.

Ce choix s'explique par deux raisons principales : tout d'abord, les périmètres de sécurité en France (*CGD* et *CSP*) ont été mis à jour depuis 2011, impliquant des ajouts, regroupements, divisions, etc. Ensuite, la politique sécuritaire en France est décidée au niveau central avant d'être appliquée dans un premier temps au niveau des 101 départements, puis dans un deuxième temps au niveau des 1200 périmètres de sécurité locaux et régionaux (*CGD* et *CSP*).

Étant donné que le découpage du territoire français en 101 départements n'a pas changé depuis 2011, nous avons choisi de regrouper les statistiques par département. Enfin, ces données seront étiquetées en fonction de la survenue ou non d'une émeute, d'un attentat, d'un attentat déjoué ou d'aucun de ces événements.

3 Apprentissage interprétable des comportements criminels

Ainsi, l'objectif de ce travail est d'appréhender l'évolution du comportement criminel à partir des données étiquetées. Nous proposons d'apprendre à partir des données correspondant à chaque année afin de déduire un comportement annuel distinct. En d'autres termes, nous développerons 10 modèles distincts, chacun correspondant à une année de 2012 à 2021. Notre approche d'apprentissage sera symbolique, reposant sur un arbre de décision, un générateur de règles d'induction/décision ou l'Analyse de Concepts Formels.

Cette méthodologie nous offre la possibilité d'interpréter et d'expliquer les modèles générés ultérieurement. Pour chaque année, nous obtiendrons un modèle d'apprentissage distinct. L'analyse de ces 10 modèles nous permettra de mieux comprendre l'évolution de la criminalité et des menaces telles que l'agitation, les émeutes, les attentats terroristes, et autres. Cette analyse consistera à comparer les arbres de décision ou les ensembles de règles générés, deux par deux. Il convient de noter que cette approche a fait ses preuves dans des travaux antérieurs (Meddouri et al., 2022).

Références

- Da Silva, A. R. C., I. C. de Paula Júnior, T. L. C. da Silva, J. A. F. de Macêdo, et W. C. P. Silva (2020). Prediction of crime location in a brazilian city using regression techniques. In *Proceeding of IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI'20)*, pp. 331–336.
- Meddouri, N., F. Rioult, et B. Cremilleux (2022). Apprentissage de comportements à partir de données temporelles hétérogènes. In *Proceeding of the EGC workshop Gestion et Analyse de données Spatiales et Temporelles with the 22^{ème} Journées Francophones en Extraction et Gestion des Connaissances (GAST/EGC'22)*.
- Tolan, G., T. Abou-El-Enien, et M. Khorshid (2015). Hybrid classification algorithms for terrorism prediction in middle east and north africa. *International Journal of Emerging Trends Technology in Computer Science (IJETTCS)* 4, 23–29.

Summary

L'intelligence artificielle est un outils incontournable de nos jours pour analyser des données et apprendre des comportements. Une analyse spatio-temporelle des données de criminalité a suscité pas mal de travaux durant la dernière décennie. Malgré la disponibilité en libre accès des données de criminalité en France sur la période de 2011 jusqu'à 2021, aucun travail de recherche n'a traité ces données, ni considéré cette problématique. Dans ce papier, nous proposons un pré-traitement de ces données (de 2012 jusqu'à 2021) pour une analyse prédictive et un apprentissage interprétable des comportements criminels en vu d'une prévention des émeutes ou des attaques terroristes.