Identifying High-risk Areas for Sexual Crimes in London using Machine Learning

# Introduction

This study aims to find a machine learning model which perform better in predicting sexual crime risk across London, while identifying key socioeconomic factors that influence high-risk area distributions.

# Literature review

1. [Yue, H. and Chen, J. (2025) ‘Interpretable spatial machine learning for understanding spatial heterogeneity in factors affecting street theft crime’, Applied Geography, 175, p. 103503. doi: 10.1016/j.apgeog.2024.103503.](https://www.sciencedirect.com/science/article/abs/pii/S0198971522000333?via%3Dihub)

This paper uses XGBoost machine learning algorithm combined with SHAP interpretation model to predict crime rate, and the results show that the proportion of non-local population and aging group contribute the most to crime prediction.

1. [González-Prieto, Á. et al. (2023) ‘Hybrid machine learning methods for risk assessment in gender-based crime’, Knowledge-Based Systems, 260, p. 110130. doi: 10.1016/j.knosys.2022.110130.](https://www.sciencedirect.com/science/article/pii/S0950705122012266)

This research uses the data of the Spanish official VioGen system and propose a mixed machine learning model that integrates Nearest Centroid and statistical methods to predict the risk of recidivism in gender-based violence cases. The results show that the hybrid model can improve the effectiveness of police protection by up to 25% comparing with existing risk assessment methods.

Research question

Data

Method

EDA

Results