

# Modeling Crime Dynamics: Predicting Burglaries in Canton Zug

\*

Luca Persia <sup>†</sup> Eduardas Lazebnyj <sup>‡</sup> Andrea Günster <sup>§</sup> Damian Kozbur<sup>¶</sup> Jérémie Decerle <sup>||</sup>

May 6, 2025

## Preliminary Version

– please do not quote without permission of the authors –

### Abstract

Predictive policing has gained considerable attention in recent years, with law enforcement agencies increasingly relying on data-driven models to anticipate and prevent crime. This study evaluates existing crime prediction methodologies and proposes an enhanced machine learning framework for predicting residential burglaries in low-crime-density areas, with a specific focus on the Canton of Zug from 2012 to 2021. We examine the limitations of traditional models based on the near-repeat principle, which struggle in regions with dispersed crime patterns. To address these challenges, we integrate ensemble modeling, feature selection, and imbalance-aware learning techniques to improve predictive accuracy. Additionally, we introduce a preliminary threshold optimization approach that dynamically adjusts classification thresholds based on population density, aiming to improve model performance in scenarios of extreme class imbalance. Our results demonstrate an improvement in identifying high-risk burglary hotspots, with Random Forest achieving the best trade-off between hit rate and false positive rate. These findings highlight the potential of advanced machine learning models in refining crime prevention strategies and optimizing law enforcement resource allocation, particularly in low-crime environments.

---

\*We sincerely thank Raphael Arnold, Dominik Balogh, Nicole Bellert, Dominik Boos, Nadine Marti, Roman Meyer, Maria Pelli, Giulio Fischietti, Alessio Schiavo, and Felix Wullschleger for their valuable feedback and insights. Data was kindly provided by the Cantonal Police Bureau (Kantonspolizei (KAPO)) Zug. We are also grateful to the participants of the *MSc Management and Law International Conference* in Luzern (2025), with special thanks to Peter Münch and Arnaud Raynouard for their thoughtful discussions. This research was supported by Innosuisse within the grant *Quantifying Illegal Activity: Estimating Dark Rates and Predicting Offenses* (<https://www.zhaw.ch/de/forschung/projekt/73547>).

<sup>†</sup>Luca Persia (University of Applied Science Zürich (ZHAW), Institute of Business Information Technology, Theaterstrasse 17, 8401 Winterthur, Switzerland; pess@zhaw.ch).

<sup>‡</sup>Eduardas Lazebnyj (University of Applied Science Zürich (ZHAW), Institute of Business Information Technology, Theaterstrasse 17, 8401 Winterthur, Switzerland; laze@zhaw.ch).

<sup>§</sup>Andrea Günster (University of Applied Science Zürich (ZHAW), Institute of Business Information Technology, Theaterstrasse 17, 8400 Winterthur, Switzerland; gues@zhaw.ch).

<sup>¶</sup>Damian Kozbur (University of Zürich (UZH), Department of Economics, Schönberggasse 1, 8001 Zurich, damian.kozbur@econ.uzh.ch).

<sup>||</sup>Jérémie Decerle (LogObject AG, Thurgauerstrasse 101 A, 8152 Opfikon, Switzerland; jeremy.decerle@logobject.ch).

# 1 Introduction

Accurately predicting criminal activity, particularly burglaries, has long been a focus of crime analysts and law enforcement agencies. This study aims to develop and validate a predictive framework for burglaries in low-crime density regions, with a specific application to the Canton of Zug in Switzerland. The goal is to generalize the approach for other areas with similar crime density. We use a dataset spanning 10 years that incorporates spatial, temporal, and sociodemographic features to construct models that identify burglary hotspots while balancing high detection rates with low false alarms.

Our approach begins with a comprehensive feature selection process. We pay particular attention to the inclusion of the near-repeat effect, a well-studied concept that suggests burglaries tend to cluster spatially and temporally. Feature selection is followed by model implementation using the methodology of [Kadar et al. \[2019\]](#). We evaluate logistic regression with L1 and L2 regularization, Random Forests, AdaBoost, and a hyper-ensemble of these models, addressing the dataset’s high imbalance through resampling and ensembling techniques. In addition, we introduce a preliminary threshold optimization technique based on population density clusters. This moving-threshold approach dynamically adjusts classification thresholds based on spatial variations, with the aim of improving prediction accuracy relative to static thresholds. However, this approach remains an initial attempt and requires further refinement to fully assess its robustness across different spatial and temporal conditions.

Model evaluation is conducted using standard Receiver Operating Characteristic (ROC) curves that provide a direct comparison of hit rates and false positive rates across varying thresholds. Our findings indicate that the near-repeat effect is not statistically significant in low-crime regions like Zug; however, its inclusion improves overall predictive performance in both static and moving-threshold models. Among the tested models, Random Forest consistently achieves the best balance between hit rate and false positive rate.

The remainder of the paper is organized as follows: Section 2 reviews related research on crime prediction, particularly burglary forecasting. Section 3 outlines our empirical strategy. Section 4 provides an overview of the dataset, while Section 5 presents the model implementation and results. Finally, Section 6 discusses key findings, limitations, and directions for future research.

## 2 Literature Review

Poisson processes are widely used in modeling crime to describe event counts over time and space. However, they assume crimes occur independently, which contradicts the observed temporal clustering of burglaries, where repeat victimization is likely ([Short et al. \[2009\]](#)). One way to address this is by adapting crime intensity predictions over time. [Taddy \[2010\]](#) introduces a Bayesian semi parametric spatial Poisson model. This approach uses dynamic linear models to adjust predictions based on evolving crime patterns, improving hotspot detection in fluctuating urban environments and making it more effective for real-time crime prevention.

Self-exciting point processes extend Poisson-based models by splitting the intensity function into a background component, representing baseline crime rates, and a foreground component, capturing event clusters. The foreground structure models crime series, similar to aftershocks in earthquakes, where one offense increases the likelihood of subsequent ones. In burglary prediction, this abstraction aligns with near-repeat patterns ([Mohler et al. \[2011\]](#), [Reinhart \[2018\]](#)). Advancements include the use of Green’s function to model long-term

cascading effects ([Kajita and Kajita \[2020\]](#)) and semi-parametric Hawkes-type models that integrate daily and weekly crime rhythms ([Zhuang and Mateu \[2019\]](#)), improving predictive accuracy in dynamic urban settings.

Gaussian and stochastic models provide probabilistic modeling of crime prediction, capturing uncertainty in spatial and temporal crime trends. [Shirota and Gelfand \[2017\]](#) propose a Log Gaussian Cox Process (LGCP) to model crime intensity in San Francisco, incorporating spatial and temporal covariates. [Flaxman et al. \[2019\]](#) enhance crime hotspot forecasting by combining Gaussian processes with autoregressive terms, improving predictions in sparse datasets. Similarly, [Morimoto et al. \[2019\]](#) use Gaussian processes to model crime propagation, demonstrating high accuracy across multiple urban datasets. While these models are flexible and scalable, their complexity makes them less practical for real-time crime prediction.

Machine learning, particularly deep learning, has advanced crime prediction by capturing complex trends in large datasets. However, these methods struggle with imbalanced or highly sparse data. [Kadar et al. \[2019\]](#) address this challenge in Switzerland by improving hit ratios through under-sampling and ensemble learning, making crime hotspot prediction more effective in low-crime-density areas. Meanwhile, [Solomon et al. \[2022\]](#) develop DeePrison, a deep learning model that integrates socio-demographic, weather, and historical crime data, using Gated Recurrent Units (GRUs) and attention layers to model spatial and temporal dependencies. While deep learning models perform well in densely populated metropolitan areas, their effectiveness in low-crime or dispersed regions remains a challenge.

Hotspot analysis is a widely used method for identifying geographic crime clusters. [Butt et al. \[2021\]](#) enhance this approach by combining Hierarchical Density-Based Spatial Clustering (HDBSCAN) with Seasonal Auto-Regressive Integrated Moving Average (SARIMA) to predict future crime incidents, improving hotspot detection accuracy. [Hajela et al. \[2020\]](#) apply KMeans clustering to integrate geographical patterns into crime predictions, demonstrating that spatial clustering methods significantly outperform models that do not leverage location data, reinforcing the importance of spatial analysis in crime forecasting.

The occurrence of burglaries depends on environmental, social, and economic factors. Demographics and socioeconomic conditions reveal who lives in a neighborhood and whether crime comes from within or outside. Local institutions and amenities can attract or deter crime. Temporal factors, like time of day, seasonality, and human routines, also create or limit opportunities for crime. Understanding these dynamics helps refine and optimize the feature set for analysis.

The near-repeat principle describes how some burglaries occur in spatial and temporal clusters, often forming crime series committed by the same offender. Empirical studies suggest that burglaries within 500 meters and 7 days of an initial incident are more likely to be linked, as offenders exploit familiar locations and repeat successful patterns ([Townsley et al. \[2003\]](#), [Johnson et al. \[2007\]](#)). These near-repeat offenders, often more methodical and professional, follow discernible patterns, making them more predictable. While near-repeat models perform well in high-crime urban environments ([Gerstner \[2023\]](#)), their predictive power diminishes in low-crime regions, where greater spatial dispersion weakens clustering effects ([Bowers and Johnson \[2004\]](#)).

Socioeconomic and demographic factors often indicate a neighborhood's burglary risk. High-income areas attract offenders from outside due to higher potential payouts, though stronger security infrastructure may act as a deterrent. In contrast, neighborhoods with high ethnic heterogeneity tend to experience more crime, likely due to lower community cohesion ([Rountree and Land \[2000\]](#), [Wang et al. \[2013\]](#)). Economic conditions also play a role—unemployment, a high ratio of rental properties to owned homes, and high vacancy rates are strong indicators of vulnerability. Transitional areas with younger populations (ages 15–24) are more susceptible to

burglary due to social disorder ([Yang \[2006\]](#)). Housing morphology further influences crime patterns: wealthier homes are more likely to be targeted during the day when unoccupied, while lower-income homes with weaker security face greater risks at night ([Coupe and Blake \[2006\]](#)).

Crime Pattern Theory suggests that urban infrastructure, landmarks, and environmental features influence where crimes occur ([Brantingham and Brantingham \[1995\]](#)). Areas with high accessibility, such as those near major roads and public transport hubs, make escape easier for offenders, increasing crime risk ([Yang \[2006\]](#)). Commercial zones—especially those with restaurants, bars, and supermarkets—offer anonymity and serve as crime attractors, while police stations and emergency services act as deterrents ([Sypion-Dutkowska and Leitner \[2017\]](#)). Urban planning also plays a role: dense residential areas with poor security face higher burglary rates due to limited surveillance and weaker social cohesion ([Yue et al. \[2017\]](#)).

Temporal and weather factors influence when burglaries happen. Low visibility can help offenders, but extreme weather, such as heavy rain or cold, can make committing a crime more difficult. Burglaries peak in the late afternoon and evening, especially on Fridays and Saturdays, following routine activity patterns ([Cohn and Rotton \[2000\]](#)). Seasonal effects matter—burglary rates are higher in summer, when people travel more and leave homes unoccupied [Calder \[2022\]](#). Public holidays affect crime differently: major holidays tend to lower burglary rates, while school closures and minor holidays create more opportunities [Cohn and Rotton \[2003\]](#). Weather also plays a role—moderate temperatures and longer daylight hours increase crime, while extreme cold or harsh conditions reduce it [Peng et al. \[2011\]](#). Recognizing these patterns helps improve predictive models by linking crime likelihood to human activity and environmental conditions.

### 3 Empirical Strategy

[Kadar et al. \[2019\]](#) address the challenge of crime prediction in low-population-density areas, such as the Canton of Aargau, where crime events are sparse and imbalanced. Traditional machine learning models struggle with imbalanced datasets. [Kadar et al. \[2019\]](#)’s hyper-ensemble framework combines multiple models with data subsampling techniques, leveraging the synergies between them to enhance crime hotspot predictions. The model integrates spatio-temporal, socio-economic, and environmental data, as well as dynamic features like weather and public events, to capture both spatial and temporal crime patterns. What sets this method apart is its combination of criminological insights, such as repeat victimization and crime pattern theory, with advanced machine learning techniques. Importantly, near-repeat cases are treated as features rather than outcomes.

#### 3.1 Addressing Class Imbalance

A key challenge in the model proposed by [Kadar et al. \[2019\]](#) is handling the extreme imbalance between crime and non-crime events. Traditional models struggle in this setting, as they often overpredict the majority class (non-crime), resulting in low precision when predicting actual crime events. Let  $N_{\text{cr}}$  represent the number of crime events and  $N_{\text{nc}}$  the number of non-crime events. The class imbalance ratio is defined as:

$$\text{Imbalance Ratio} = \frac{N_{\text{cr}}}{N_{\text{cr}} + N_{\text{nc}}}$$

In [Kadar et al. \[2019\]](#)’s dataset, this ratio is approximately 0.06%, indicating that non-crime events outnumber crime events by a factor of roughly 1700. Preliminary tests show that the imbalance ratio can exceed

Kadar et al. [2019]’s estimates on certain days, suggesting an even more severe potential class imbalance. Kadar et al. [2019] propose an imbalance-aware hyper-ensemble, which combines two powerful techniques:

1. **Random Under-Sampling:** This technique creates balanced subsamples of the data by randomly selecting a subset of the majority class (non-crime events) that matches the size of the minority class (crime events). This helps ensure that the model does not disproportionately focus on non-crime data.

Let  $D$  be the full dataset, and  $D_{\text{nc}}$  and  $D_{\text{cr}}$  be the majority (non-crime) and minority (crime) classes, respectively. Random undersampling draws a subset  $D'_{\text{nc}} \subseteq D_{\text{nc}}$  such that:

$$|D'_{\text{nc}}| = |D_{\text{cr}}|$$

This creates a balanced training set  $D' \subseteq D$ .

2. **Ensemble Learning:** Multiple models are trained on different subsamples, and their predictions are averaged. This approach leverages the strengths of each model and mitigates the potential biases that can result from any single model.

For an ensemble of models  $f_1, f_2, \dots, f_M$ , the final prediction  $\hat{y}$  is the averaged output:

$$\hat{y} = \frac{1}{M} \sum_{m=1}^M f_m(x)$$

where  $x$  is the input feature vector, and  $f_m$  is the prediction of the  $m$ -th model.

Experiments show the best results come from an ensemble of three base models:

- **Random Forests:** This method builds multiple decision trees on random subsets of features and data, aggregating their predictions to reduce overfitting and increase stability. It is particularly effective with large, high-dimensional datasets containing noisy features.
- **AdaBoost (Adaptive Boosting):** AdaBoost sequentially trains classifiers, focusing on misclassified instances from previous models, and combines weak learners into a strong classifier. It minimizes both bias and variance, particularly in datasets with complex patterns.
- **Logistic Regression with L1/L2 Regularization:** This linear classification model uses a logistic function to estimate the probability of an input belonging to a certain class. L1 regularization encourages sparsity for feature selection, while L2 regularization helps prevent overfitting by penalizing large coefficients.

Combining both techniques can be done in a straightforward and synergetic way. This approach allows us to retain the best characteristics of each method, leading to significantly better performance. As a result, the hyper-ensemble achieves substantial improvements, especially when predicting the top 5% and 20% of crime hotspots, with the hit ratio increasing from 18.1% to 24.6% for the top 5% and from 53.1% to 60.4% for the top 20%. This approach, first applied in the canton of Aargau, demonstrates scalability to other cantons, where crime incidents are similarly infrequent and dispersed across large geographic areas.

## 3.2 Potential Model Improvements

Kadar et al. [2019] include several well-designed elements significantly enhancing crime prediction, though further considerations may be useful.

### 3.2.1 Handling of Class Imbalance

In our study, we considered different resampling methodologies to address the problem of class imbalance in the dataset. Below, we outline the challenges associated with commonly used approaches and justify our current choice. Random Oversampling involves random duplication of minority class events to balance the dataset. However, this approach can introduce significant bias, especially in a panel dataset where temporal consistency is critical. For example, random duplication of events could result in multiple identical events occurring on the same day and at the same location, thus misrepresenting the temporal structure of the data. The same applies to more statistically robust oversampling techniques such as SMOTE [Chawla et al. 2002]. It generates synthetic samples for the minority class by interpolating among existing observations rather than simply duplicating or discarding data. But likewise, generating synthetic samples can distort the temporal and spatial structure of the data. In addition, SMOTE is computationally expensive, particularly for large datasets with high dimensionality, making it less practical for our use case.

In light of these considerations, we opted to stick with Random Undersampling, in which the majority class observations are dropped randomly to achieve class-balance. An important problem is that undersampling may introduce selection bias by potentially discarding informative observations from the majority class. But it turns out to be the best trade-off between maintaining temporal consistency and computational efficiency, as it actually reduces the size of the dataset and, from a deployment perspective, makes it faster to return reliable results.

One approach to be further investigated involves cost-sensitive learning, which assigns higher penalties for misclassifying crime events, thereby improving the model’s handling of imbalanced datasets. Consider the equation:

$$\mathcal{L}(y, \hat{y}) = w_{cr} \cdot \mathcal{L}_{cr}(y_{cr}, \hat{y}_{cr}) + w_{nc} \cdot \mathcal{L}_{nc}(y_{nc}, \hat{y}_{nc})$$

where  $w_{cr}$  and  $w_{nc}$  are the weights assigned to crime and non-crime losses, and  $\mathcal{L}$  is a loss function such as cross-entropy or mean squared error. Increasing  $w_{cr}$  would prioritize the accurate classification of crime events, even though they are less frequent.

### 3.2.2 Granularity of Predictions

The model proposed by Kadar et al. [2019] is trained on a spatiotemporal dataset divided into 200m x 200m grid cells. The choice of grid size plays a fundamental role in addressing class imbalance, ensuring that rare events are not excessively sparse, while simultaneously improving the practical applicability of the model for law enforcement by providing spatially accurate predictions. In our setup, we provide predictions for a more granular grid system of 100m x 100m.

### 3.2.3 Testing in High-Density Areas

Kadar et al. [2019]’s model is tailored for low-density areas with sparse crime events. Testing the hyper-ensemble in high-density regions, where crime events are more frequent and concentrated, would provide insights into its scalability. This could also involve adjusting the model to account for higher levels of heterogeneity in urban areas, where different socio-economic and geographic features play a more prominent role.

## 4 Burglaries in Canton Zug - Descriptive Statistics

We use the Kadar model as a benchmark for crime prediction in the canton of Zug, adapting it to better suit low-density areas. While we follow Kadar’s general methodology, we introduce key modifications to improve accuracy and relevance for our specific setting.

### 4.1 RIPOL Code Matching: Identifying Burglary

We derive burglaries from the population of all recorded offenses, where the police records various details for each incident, including victims, location, modus operandi, stolen items, and devices used. One critical piece of information is the legal Article of the Swiss Criminal Code (SCC) violated by the perpetrator. While SSC Art. 139 broadly defines burglary, it also covers cases such as car theft (Grand Theft Auto) and other offenses. Additionally, burglary can be a secondary element in more severe crimes, such as murder or rape, which are outside the scope of this study.

To accurately define residential burglary, we implement a RIPOL (automated system for Police search) code-matching approach. This method systematically identifies burglary-related incidents by filtering for combinations of specific legal articles, following classification procedures established by forensic investigators at the cantonal police bureau (Kantonspolizei (KAPO)) Zug. Table 1 provides examples of how legal articles are combined to classify burglary incidents.

Incident Type	Article 1	Incident Type	Article 2
Burglary	SSC Art. 139	Trespassing	SSC Art. 186
Burglary	SSC Art. 139	Damage to Property	SSC Art. 144 Abs. 1

Table 1: Examples of Article Combinations Identifying Burglary

We classify incidents based on the presence of two RIPOL codes recorded, evaluating different permutations of these codes to categorize cases. To define residential burglary, we include all incidents that fall under the following three classifications:

- Illegal entry with intent to commit burglary
- Incidents primarily involving property damage or unauthorized entry
- Opportunistic or misappropriative thefts and minor property crimes

These classifications ensure that only relevant burglary cases are retained, filtering out incidents that do not align with our focus.

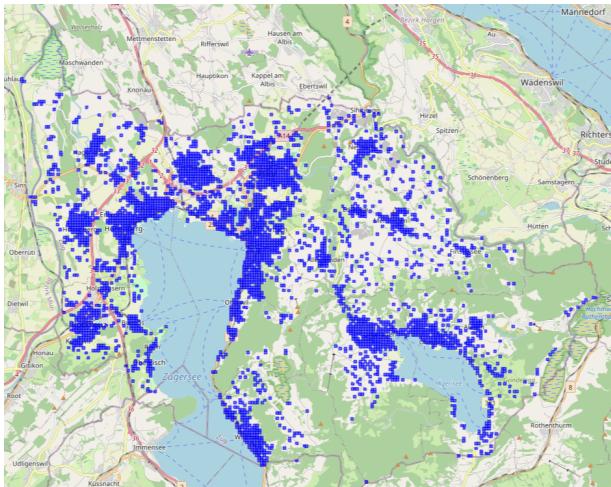
## 4.2 Temporal Filtering of Incidents

In many cases, burglaries cannot be assigned an exact time, as they fall within a time window. Records include both a start and an end date for each incident, with the end date sometimes extending one to two months after the start. This occurs when property owners report burglaries long after they happen, often due to prolonged absences (e.g., vacations, working/staying abroad) or because the residence is not regularly occupied. To improve the temporal accuracy of our dataset, we restrict incidents to those with a duration between 0 and 48 hours. This prevents long, uncertain time spans from affecting the analysis of crime patterns over time.

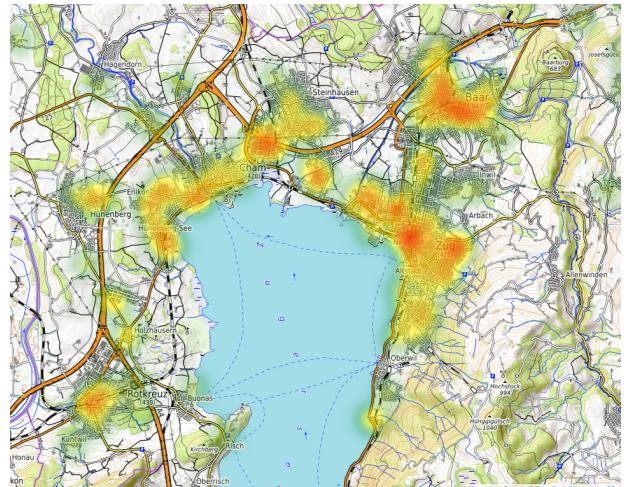
## 4.3 Overview of Burglaries in Canton Zug

We analyze daily crime data from January 1, 2012, to December 15, 2021, incorporating temporal, spatial, and demographic features. To structure the data, we implement a  $100m \times 100m$  grid system, maintaining the granularity provided by the Swiss Federal Statistical Office Swiss rather than aggregating cells and thereby features (Kadar et al. [2019]). The time horizon we discuss here represents the full sample, which we use for predicting burglaries past 2021.

We filter the data to only include burglary incidents based on legal Article combinations. The final dataset contains 12,884,455 observations distributed across 3,789 grid cells (Figure 1a). Of the total observations, only 2,696 are burglaries, highlighting a stronger class imbalance compared to Kadar et al. [2019] due to the more granular grid size. The final burglary rate is 0.0209% for the Canton of Zug. While we have information for the entire Canton Zug, we only consider grids including houses, excluding breakages of cars, trucks etc. We aim to predict the next cell in which a burglary will happen for a daily time horizon. Figure 1a indicates in blue the cells relevant for burglary prediction.



(a) Zug as a Grid ( $100m^2$ )



(b) Heatmap of Burglaries in 2021

Figure 1: Canton Zug: Grid and Heatmap of Burglaries

Figure 1b presents the heatmap of actual burglaries in Canton Zug for the year 2021. In Appendix B, Figure 6 illustrates temporal crime trends, showing heatmaps for all years from 2012 to 2021. We use the training set to train the models for prediction. Figure 1a and Figure 1b show that burglaries mostly happen in the highly populated areas of Canton Zug, like its capital, the city of Zug, and cities along the shore of Lake Zug.

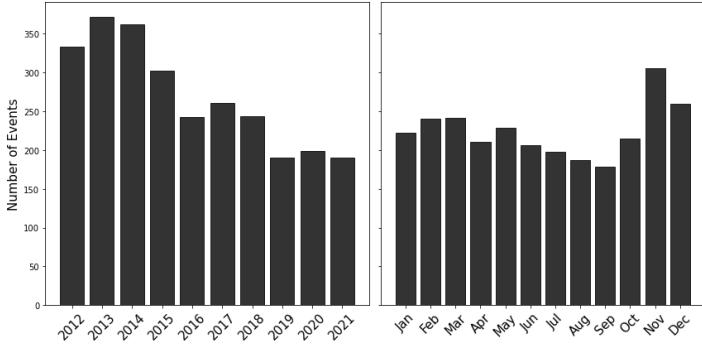


Figure 2: Number of Burglaries per Year (Left Panel) and Month (Right Panel) (2012-2021)

Figure 2 shows the count of burglaries taking place per year (left panel) and per month (right panel). The left panel indicates that burglaries are decreasing over our sample horizon. The right panel confirms the finding in the literature; burglaries mainly take place in the winter months with early sunsets.

## 4.4 Feature Engineering

To predict burglaries, we collect a set of 45 features of which 25 are continuous, 11 are categorical and 9 are binary. The choice of variables is based on the literature aiming to predict criminal hotspots, particularly burglaries. Kadar et al. [2019] also analyzes a Swiss canton (Aargau), we follow to a large extent her feature choice set. Given the high-dimensionality of the dataset and the (potential) high computational power required to run a machine learning model, we perform a set of different steps for feature scaling and selection. Appendix A states all features we use, including their source, frequency, and dimension.

Figure 3 shows the population density by event class. Median population density is significantly higher for areas experiencing events ( $flag = 1$ ), slightly above 50 people per hectare. Our analysis confirms population density to be the key factor influencing burglary occurrence. This aspect requires particular attention when training our model. It might potentially lead to over-predict high population density areas. Note also that population density stays fairly constant over time, not only because of it being annually updated. Beyond population density, we also consider the demographic diversity of the population within each cell, including factors such as age distribution, ethnic diversity, and other socio-economic characteristics. Table 7 in Appendix A provides a detailed breakdown.

In Figure ?? we compare the average daily temperature and daily precipitation levels (in millimeters) for burglary events and non-events. Data points for events are shown with distinct markers to differentiate them from non-events. The graph reveals the importance of weather, indicating a potential correlation between low precipitation levels and the preponderance of events, as highlighted by the literature on occurrence of burglaries. This also suggests that integrating a more granular collection of weather-related atmospheric and environmental data into the dataset could reveal additional patterns and potential correlations.

### 4.4.1 Preprocessing

Numeric features are standardized by removing the mean and scaling to unit variance. This ensures that all features are on a comparable scale while preserving the relative differences in the data. Binary and categorical features are preprocessed separately. We treat as binary all features with two possible states, such as true/false, yes/no, or 1/0, and as categorical all features with discrete categories, such as region names, years and months,

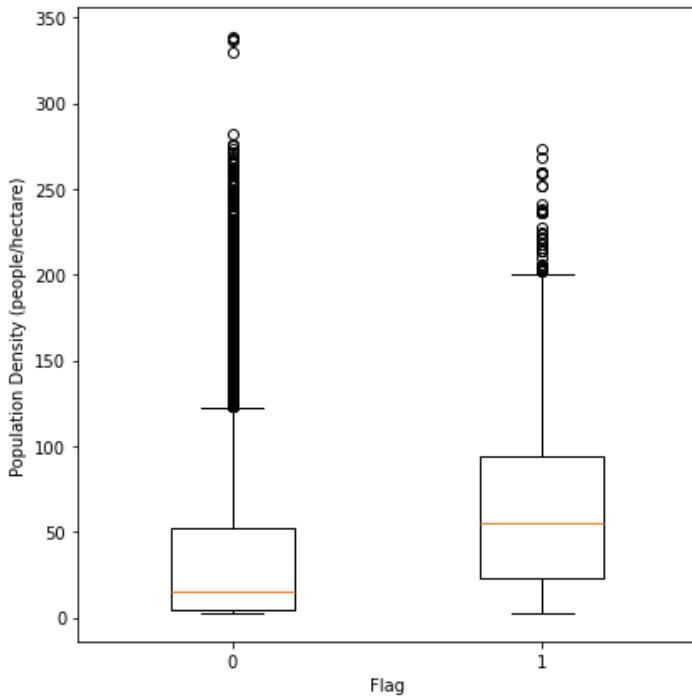


Figure 3: Population Density by Event ( $Flag = 1$  indicating Burglary) (2012-2021)

or number of infrastructures in a grid. For binary features, the only operation to be performed is to ensure that missing items are replaced by the most frequent value for that feature.

Categorical features require more comprehensive preprocessing steps to convert their non-numeric nature into a format that can be used by machine learning or statistical algorithms. First, missing values are replaced with the most frequent category for each feature, as for binary features. Then, one-hot encoding is used to ensure that each category within a feature is represented as a separate binary column. This encoding guarantees that there are no ordinal relationships between categories.

## 5 Predicting Burglaries in Canton Zug - Results

In this section we perform feature selection by combining a logistic regression with a backward elimination algorithm. We then implement the model proposed by Kadar et al. [2019] and introduce a novel approach, a *moving-threshold model* which optimizes the classification threshold using a clustering rule and a trade-off between specific metrics. To evaluate our results, we do not rely on *precision*, which measures the proportion of correctly predicted positive events out of all predicted positives. While precision and recall are commonly optimized together in imbalanced classification tasks using metrics such as the *F1-Score* or through analysis of the Precision-Recall Curve, this approach is not well-suited for our case due to the extreme imbalance and large size of our dataset. The potential improvement in precision would be minimal and difficult to measure accurately, making it challenging to distinguish between models.

Instead, we focus on two key metrics: *hit-rate* (HR), which is equivalent to recall or to the True Positive Rate (TPR) and measures the proportion of correctly detected burglaries, and the *False Positive Rate* (FPR), which measures the proportion of false alarms among the non-event locations. These metrics directly address the practical trade-offs in crime prediction: HR emphasizes the model’s ability to correctly detect burglary events, while FPR captures the risk of over-responding to non-burglary events. To evaluate the models, we

use the Receiver Operating Characteristic (ROC) curve, which plots the HR against the FPR. Good models are close to the top-left corner, indicating a high hit rate and a low false positive rate, while poor models will exhibit a more linear relationship, indicating little distinction between events and non-events.

The ROC curve is especially valuable in two contexts (i) imbalanced classification tasks with highly imbalanced datasets and (ii) real-world applications where both under-detection (missed events) and over-detection (false alarms) carry significant consequences, such as in crime prevention or disease control.

## 5.1 Feature Selection

To perform feature selection, we follow a two-step procedure. In the first step, we perform a correlation analysis, calculating both Pearson and Spearman correlation coefficients for all continuous and binary features, and we drop the features that could raise multicollinearity issues. The second step involves specifying a logit model for the remaining features, incorporating the near-repeat effect variable only for a 7-day window and within a 500-meter radius. The categorical features were retained without being included in the feature selection process. This decision was based on the established literature regarding points of interest and the importance of maintaining temporal effects through variables such as *year* and *month*.

For the correlation analysis, Pearson's and Spearman's correlation coefficients were calculated to better understand the relationships between features. Pearson's correlation is calculated based on the actual values of the features and is suitable for capturing linear relationships. In contrast, Spearman's correlation is calculated on the ranks of the data and is more suitable for detecting monotonic relationships, which may not be linear. We find this dual analysis particularly useful in our case, as we assume that the relationships between the features are not strictly linear. In Figure 5, shown in the Appendix A, we report the respective correlation matrices. We drop variables that have an absolute value correlation greater than 0.80, to avoid possible multicollinearity issues. In particular, we look at the cross-sectional results of the two correlation matrices, and we filter out the feature *busidens* from the dataset.

We proceed by employing a logistic regression model, which is the classical logit model specified as:

$$\log \left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k,$$

where  $p$  represents the probability of an event occurring,  $\beta_0$  is the intercept,  $\beta_i$  represents the coefficients of the predictors  $X_i$ , and  $k$  denotes the total number of predictors.

To further refine the feature selection methodology, we employ a backward elimination algorithm. This step-wise approach iteratively removes predictors from the model based on their statistical significance, specifically using a  $p$ -value threshold of 0.05. The backward elimination algorithm is summarized as follows:

**Dep. Variable:** flag    **burglary=1**    **no-burglary=0**  
**No. Observations:** 12884455  
**Model:** Logit  
**Method:** MLE  
**Pseudo R-squ.:** 0.05575

	coef	std err	z	P> z	[0.025	0.975]
<b>Intercept</b>	-7.1264	0.304	-23.413	0.000	-7.723	-6.530
<b>tavg</b>	-0.8120	0.108	-7.524	0.000	-1.023	-0.600
<b>prcp</b>	-0.9520	0.265	-3.599	0.000	-1.471	-0.434
<b>popdens</b>	0.0030	0.001	4.892	0.000	0.002	0.004
<b>popstab</b>	-0.7628	0.175	-4.355	0.000	-1.106	-0.419
<b>popcit_CH</b>	-0.6721	0.249	-2.698	0.007	-1.160	-0.184
<b>popcit_EU</b>	-0.9844	0.261	-3.773	0.000	-1.496	-0.473
<b>popcit_europe</b>	-0.7154	0.322	-2.220	0.026	-1.347	-0.084
<b>popcit_noneurope</b>	-1.1377	0.318	-3.578	0.000	-1.761	-0.514
<b>popcit_other</b>	5.2863	1.957	2.701	0.007	1.450	9.123
<b>pop_age1</b>	-0.5464	0.108	-5.082	0.000	-0.757	-0.336
<b>pop_age2</b>	-0.1974	0.091	-2.178	0.029	-0.375	-0.020
<b>pop_age3</b>	-0.4090	0.070	-5.850	0.000	-0.546	-0.272
<b>pop_age4</b>	-0.3447	0.083	-4.158	0.000	-0.507	-0.182
<b>popcit_dividx</b>	0.5676	0.116	4.891	0.000	0.340	0.795
<b>popage_dividx</b>	0.4060	0.131	3.093	0.002	0.149	0.663
<b>busisec1</b>	-0.8917	0.152	-5.852	0.000	-1.190	-0.593
<b>busisec2</b>	-0.6635	0.110	-6.006	0.000	-0.880	-0.447
<b>busisec3</b>	0.6781	0.074	9.191	0.000	0.533	0.823
<b>busisec_dividx</b>	1.0058	0.087	11.540	0.000	0.835	1.177
<b>empldens</b>	0.0019	0.000	15.766	0.000	0.002	0.002
<b>emplmale</b>	-0.1927	0.063	-3.046	0.002	-0.317	-0.069
<b>emplsec_dividx</b>	-0.4267	0.105	-4.068	0.000	-0.632	-0.221
<b>is_weekend</b>	-0.1287	0.039	-3.268	0.001	-0.206	-0.052

Table 2: Backward Feature Selection: Logit Summary.

---

**Algorithm 1** Backward Elimination with a Logistic Regression

---

1. **Initialize:** Fit the logistic regression model with all predictors.
  2. **Evaluate Significance:** Compute the  $p$ -value for each predictor in the model.
  3. **Eliminate Predictors:** Identify the predictor with the highest  $p$ -value (greater than 0.05) and remove it from the model.
  4. **Refit the Model:** Refit the logistic regression model with the remaining predictors.
  5. **Repeat:** Repeat steps 2–4 until all remaining predictors have  $p$ -values less than or equal to 0.05.
- 

This iterative procedure ensures that the final model retains only statistically significant predictors, facilitating its interpretation and robustness. The results of the final logistic regression model, after backward elimination, are presented in Table 2. This table provides the estimated coefficients ( $\beta_i$ ), their standard errors, and corresponding  $p$ -values for all retained predictors.

At the end of the feature selection process, we excluded the variables *popmale*, *popbirth\_non\_CH*, *busidens*, and the near-repeat feature (*nr\_500\_7*) from the analysis. The removal of *nr\_500\_7* is particularly noteworthy,

given the strong emphasis on the near-repeat effect in the existing literature. This finding is both empirically and statistically significant, as it suggests that the near-repeat effect is not observable in the Zug Canton. More broadly, this raises the possibility that rural cantons with low crime density may not exhibit strong near-repeat patterns. To further investigate this hypothesis, we conducted additional robustness checks during the model deployment phase, estimating models both with and without the near-repeat effect to validate this result.

## 5.2 Comparison to Kadar et al. [2019]

Our ensemble approach consists of four types of models: logistic regression with L1 ( $LR_{L1}$ ) and L2 ( $LR_{L2}$ ) regularization, random forests (RF) and Adabtive Boosting (AdaBoost). The ultimate goal is to avoid overfitting the majority class, which without resampling would be predicted in most (i.e., almost all) cases. Unlike the algorithm developed in Kadar et al. [2019], we added an additional step for fine-tuning the hyperparameters for each model by performing a k-fold cross-validation, a technique widely used in machine learning and clearly explained in Stone [1974].

As the training set  $X_{train}, y_{train}$  we consider seven years of observations from January 1, 2012, to January 1, 2019, while the test set  $X_{test}, y_{test}$  includes the remaining three years from January 1, 2019, to December 15, 2021. Compared with the three years (2 years training, 1 year testing) analyzed in Kadar et al. [2019], our application has a much larger training set and is able to generalize more accurately on the test set. We define the ensemble size  $\phi = 10$  and the model set  $Q = [LR_{L1}, LR_{L2}, RF, AdaBoost]$ . The parameter k for cross-validation is set equal to 5. We also define the threshold to calculate the binary outcome as  $\tau = 0.5$ . The ensemble model algorithm is summarized as follow:

---

### Algorithm 2 Ensemble Model

---

**Require:**  $X_{train}, y_{train}; X_{test}; \phi; Q; \tau$

**Training Phase:**

Initialize the ensemble  $\mathcal{E} = \emptyset$

**for**  $i = 1$  to  $\phi$  **do**

    Random Under-Sampling  $X_{train}, y_{train}$  without replacement to create a bootstrapped balanced dataset

**for**  $N \in Q$  **do**

        Train model  $M^{i,N}$  on the resampled data

        Perform k-fold cross-validation to optimize hyperparameters

        Add  $M^{i,N}$  to the ensemble  $\mathcal{E}$

**end for**

**end for**

**Prediction Phase:**

**for**  $M \in \mathcal{E}$  **do**

    Apply  $M$  to  $X_{test}$  to obtain probabilities  $\hat{y}^i$

**end for**

Compute final probabilities:  $\hat{y} = \frac{1}{\phi} \sum_{i=1}^{\phi} \hat{y}^i$

Compute the binary outcome:  $y_{pred} = \mathbb{1}_{\hat{y} > \tau}$

**return** Predicted outcomes  $y_{pred}$ ; Probabilities  $\hat{y}$ ; Metrics  $HR$  and  $FPR$ ) =0

---

We trained and tested the four models specified in  $Q$ , both individually and as part of the ensemble algorithm. The evaluation was conducted in two distinct setups: first, including the feature representing the near repeat effect (within 7 days and a radius of 500 meters), and second, excluding the near repeat feature.

Figures 4a and 4b show the ROC curves under both specification, while Tables 10 and 11 in the Appendix C provide the corresponding evaluation metrics for a set of thresholds and highlight the default threshold of 0.5. The Random Forest (RF) consistently achieves the best overall balance between HR and FPR across a range

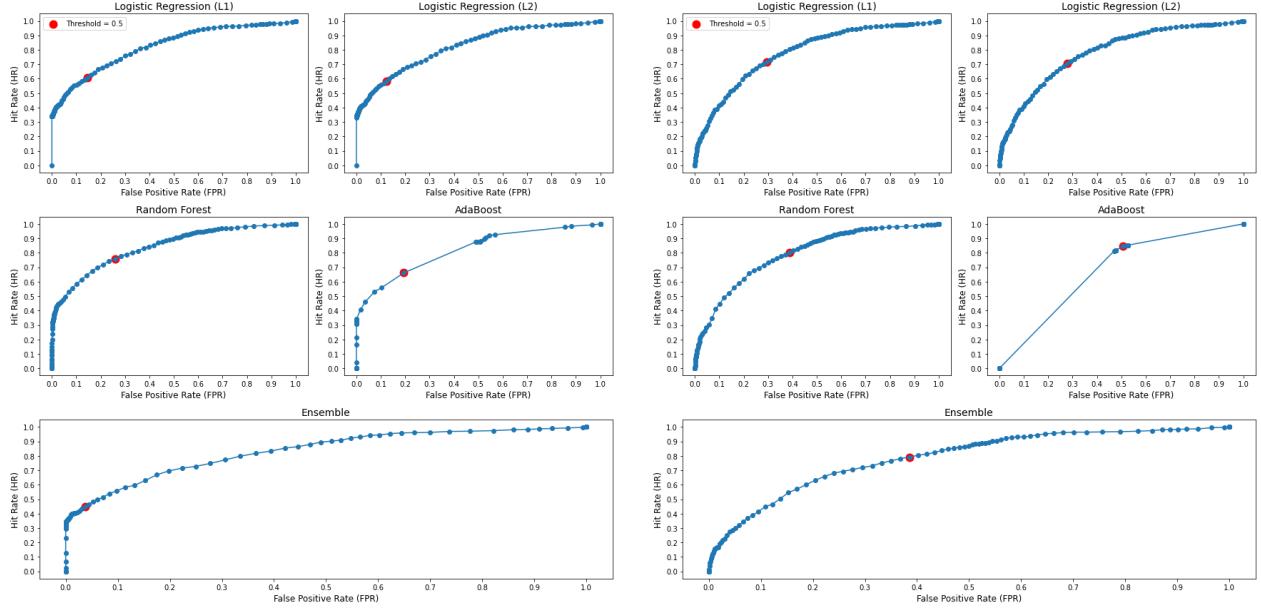


Figure 4: ROC Curves (Hit Rate and False Positive Rate).

of thresholds. RF typically maintains a high HR while controlling FPR better than the other models. The Ensemble method achieves particularly low FPR values, though at the cost of a lower HR in many cases. Logistic regression models generally provide a moderate balance between HR and FPR, but they do not outperform RF or the Ensemble. AdaBoost (AB) exhibits extreme behavior, with a high HR accompanied by high FPR at lower thresholds. As the threshold increases, both HR and FPR approach zero. This instability shows sensitivity to noise or overfitting, suggesting that the AB algorithm is not well suited for the extreme imbalance in our dataset. When we exclude the near repeat effect, the logistic regression models maintain consistent trade-offs but do not surpass the RF, which continues to offer the most balanced performance also in this case, achieving a high HR with a relatively controlled FPR. The Ensemble remains effective in minimizing FPR, but it comes at the cost of a lower HR. At a threshold of 0.5, RF achieves 75.64% HR and 25.85% FPR. Additionally, RF has significantly lower computational requirements compared to the Ensemble, a crucial aspect in terms of deployment of a potential application for the Zug Canton Police.

Overall, we show that we could aim to optimal set the threshold by using a ROC curve as a diagnostic instrument. Moreover, the inclusion of the near repeat effect did not results in a lower HR and overall worse model performance, as we would have expected from the feature selection section which identified the NR effect as statistically insignificant for the Canton of Zug.

### 5.3 Moving-Threshold Optimization

In this section, we propose a novel moving-threshold optimization approach to address potential biases in model performance caused by population density variations. As observed in Figure 3 of Section 4.4, the median population density for burglary events is significantly higher than for non-events. This indicates that areas with higher population densities are more prone to burglary events, potentially biasing the model toward over-predicting in these areas. Consequently, this can lead to an increased false positive rate for high-density areas.

To mitigate this issue, we cluster by population density and train a separate model for each cluster. We first

train a global model on the usual test set. We then use a validation set to optimize the classification threshold by maximizing a specific function of the hit rate and false positive rate, subject to a minimum hit rate. This ensures that the threshold is tailored to the specific characteristics of each population density segment. We evaluate the optimized thresholds by testing the models on a held-out test set.

As the training set  $X_{train}, y_{train}$ , we consider six years of observations from January 1, 2012, to January 1, 2018. The validation set  $X_{val}, y_{val}$  includes two years of data from January 1, 2018, to January 1, 2020, while the hold-out test set covers the remaining two years until December 15, 2021. We specify a minimum hit rate of  $HR_{min} = 50\%$ . The set of candidate thresholds is defined as  $\tau_q = [0.1, 1]$  with increments of 0.1. To segment the population density effectively, we apply a simple k-means algorithm, grouping the data into  $k = 5$  clusters based on the variable *popdens*. The choice of five clusters ensures that the model captures significant variations in population density while maintaining computational feasibility. The clustering allows us to assign an optimized threshold  $\tau_k$  to each cluster, tailored to the characteristics of the population density within that group. In order to account for the fact that the distributions of features in the training and validation sets may differ from that of the test set, for test observations where the optimized threshold  $\tau_k$  for the respective cluster results in worse performance than the best overall threshold identified in Chapter 5.2, we override the cluster-specific threshold and apply the best threshold from Chapter 5.2. This adjustment ensures robustness in the model's predictions by avoiding performance degradation due to unseen population density distributions in the test set. The algorithm is summarized as follow:

---

**Algorithm 3** Threshold Optimization

---

**Require:**  $X_{train}, y_{train}, X_{val}, y_{val}, X_{test}, y_{test}, k, \tau, HR_{min}$ 

 Train a single global model  $M$  on the undersampled  $X_{train}, y_{train}$ 
**Clustering Phase:**

 Apply k-means clustering to population density in  $X_{val}$ , assigning  $k$  clusters

**for each** cluster  $k$  **do**
**for each** candidate threshold  $\tau_k$  **do**

 Apply  $M$  to compute probabilities  $P(y = 1 | X_k)$  on the validation set

 Compute  $HR_k(\tau)$  and  $FPR_k(\tau)$ 

Evaluate an objective function:

$$\min_{\tau_k} f(FPR_k(\tau)),$$

$$\text{s.t. } HR_{min}$$

$$\tau_k \in [0, 1]$$

 Store  $\tau_k^*$  that minimizes the objective function

**end for**
**end for**
**Prediction Phase:**

 Obtain probabilities  $\hat{P}(y = 1 | X_{test})$ 

 Assign each observation in  $X_{test}$  to its corresponding population density cluster  $k$ 
**for each** observation in  $X_{test}$  **do**

 Apply the optimized threshold  $\tau_k^*$  for the corresponding cluster to make binary predictions

**end for**
**Decision Rule for Global Prediction:**
**for each** observation in  $X_{test}$  **do**

Assign the final prediction based on the following rule:

$$y_{pred} = \begin{cases} \hat{y}_{\text{moving}} & \text{if } HR_{\text{moving}} \geq HR_{\text{fixed}} \text{ and } FPR_{\text{moving}} \leq FPR_{\text{fixed}} \\ \hat{y}_{\text{fixed}} & \text{otherwise} \end{cases}$$

**end for**
**return** Predicted outcomes  $y_{pred}$ ; Probabilities  $\hat{y}$ ; Metrics  $HR$  and  $FPR$ ; Optimal Threshold  $\tau_k^* = 0$ 


---

We formulate our optimization problem as:

$$\min_{\tau_k} FPR_k(\tau_k) + \lambda \cdot \max(0, HR_{min} - HR_k(\tau_k)),$$

$$\text{s.t. } HR_{min} = 0.5$$

$$\tau_k \in [0, 1]$$

 where  $\tau_k$  is the classification threshold for cluster  $k$ ,  $FPR_k$  is the false positive rate for cluster  $k$  at threshold  $\tau_k$ ,  $HR_k$  is the hit rate for cluster  $k$  at threshold  $\tau_k$  and  $\lambda$  is a penalty term for solutions where  $HR_k < 0.5$ .

Cluster	Without NR			With NR		
	HR (%)	FPR (%)	$\tau_k^*$	HR (%)	FPR (%)	$\tau_k^*$
0	66.67	20.40	0.40	50.00	12.14	0.40
1	50.00	37.74	0.65	36.51	13.06	0.60
2	0.00	21.96	0.65	49.04	4.27	0.65
3	50.00	39.79	0.55	68.22	11.35	0.55
4	100.00	46.70	0.65	72.73	48.52	0.55
<b>Global</b>	51.20	13.50		47.80	3.10	

Table 3: Comparison of Hit Rate (HR) and False Positive Rate (FPR) with and without the Near-Repeat Effect with optimized thresholds per population cluster.

Table 3 presents the moving threshold results, highlighting the global predictions where the best threshold

from the model trained in Section 5.2 are applied in cases where the moving threshold underperforms. The overall performance is similar to the best RF model in our implementation of Kadar (Appendix C), with a 44.73% HR and a 3.69% FPR when including the NR effect, and a 51.20% HR and 13.50% FPR without the NR effect. The intra-cluster results show promising performance, particularly in some clusters - cluster 2 with the NR effect is close to a 50% HR with a FPR lower than 5% - indicating that the moving threshold provides interesting preliminary results and that a refinement of the optimization or clustering approach can potentially mitigate the imbalance problem. However, the algorithm is sensitive to differences in the distribution of population density across the training, validation, and test sets. Additionally, performance differences across clusters may result from varying burglary rates, which in some cases may be significantly lower or even zero in the test set compared to the validation set.

## 6 Discussion

In this study, we evaluated several existing statistical and machine learning methodologies for crime prediction and proposed an improved approach for predicting daily residential burglaries in the Canton of Zug. Our approach is particularly relevant for areas with low crime density.

We conducted a rigorous exploratory data analysis, including feature preprocessing and selection using logistic regression with a backward elimination algorithm. This initial analysis allowed us to assess the near-repeat effect, which, despite being widely supported in the literature, it was not statistically significant in our case. We implemented four machine learning models: two regression-based (logistic regression with L1 and L2 regularization) and two tree-based (random forest and adaptive boosting). To address the severe class imbalance — where only 0.02% of observations correspond to burglary events — we applied random undersampling without replacement.

Our current setup serves as a baseline, but further optimizations are needed to validate robustness across different threshold scenarios. This is particularly relevant for the near-repeat effect, as our results suggest that it has limited influence on burglary patterns in low-crime regions. However, including the feature in the model improved the hit rate–false positive rate (HR-FPR) ratio compared to models that excluded it.

For future research, we propose investigating self-exciting point processes, such as Hawkes processes [Reinhart 2018], to analyze crime dynamics more effectively. These models can capture delayed or indirect near-repeat effects that may not be immediately observable in classical machine learning frameworks. Additionally, point processes inherently mitigate data imbalance issues by focusing exclusively on event distributions.

## References

- Bowers, K. and Johnson, S. D. (2004). Who commits near repeats? a test of the boost explanation. *Western Criminology Review*, 5:12–24.
- Brantingham, P. and Brantingham, P. (1995). Criminality of place: Crime generators and crime attractors. *Eur J Crim Policy Res*, 3:5–26.
- Butt, U. M., Letchumanan, S., Hassan, F. H., Ali, M., Baqir, A., Koh, T. W., and Sherazi, H. H. R. (2021). Spatio-temporal crime predictions by leveraging artificial intelligence for citizens security in smart cities. *IEEE Access*, 9:47516–47529.
- Calder, J. D. (2022). *Burglary Research and Conceptualizing the Community Security Function, a Learning Organization*, pages 193–216. Springer International Publishing.
- Chawla, N., Bowyer, K., Hall, L., and Kegelmeyer, W. (2002). Smote: Synthetic minority over-sampling technique. *J. Artif. Intell. Res. (JAIR)*, 16:321–357.
- Cohn, E. G. and Rotton, J. (2000). Weather, seasonal trends and property crime in minneapolis, 1987-1988. a moderator-variable time-series analysis of routine activities. *Journal of Environmental Psychology*, 20(3):257–272.
- Cohn, E. G. and Rotton, J. (2003). Even criminals take a holiday: Instrumental and expressive crimes on major and minor holidays. *Journal of Criminal Justice*, 31(4):351–360.
- Coupe, T. and Blake, L. (2006). Daylight and darkness targetting strategies and the risks of the being seen at residential burglaries. *Criminology*, 44(2):431–464.
- Flaxman, S., Chirico, M., Pereira, P., and Loeffler, C. (2019). Scalable high-resolution forecasting of sparse spatiotemporal events with kernel methods: A winning slution to the nij "real time crime forecasting challenge". *The Annals of Applied Statistics*, 13(4):2564–2585.
- Gerstner, D. (2023). Predictive policing in the context of residential burglary: An empirical illustration on the basis of a pilot project in baden-württemberg, germany. *European Journal for Security Research*, 3:115–138.
- Hajela, G., Chawla, M., and Rasool, A. (2020). A clustering based hotspot identification approach for crime prediction. *Procedia Computer Science*, 167:1462–1470. International Conference on Computational Intelligence and Data Science.
- Johnson, S. D., Bernasco, W., Bowers, K., Eiffers, H., Ratcliffe, J. H., Rengert, G. F., and Townsley, M. K. (2007). Space-time patterns of risk: A cross national assessment of residential burglary victimization. *Journal of Quantitative Criminology*, 23.
- Kadar, C., Maculan, R., and Feuerriegel, S. (2019). Public decision support for low population density areas: An imbalance-aware hyper-ensemble for spatio-temporal crime prediction. *Decision Support Systems*, 119:107–117.
- Kajita, M. and Kajita, S. (2020). Crime prediction by data-driven green's function method. *International Journal of Forecasting*, 36(2):480–488.
- Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., and Tita, G. E. (2011). Self-exciting point process modelling of crime. *Journal of the American Statistical Association*, 106(493):100–108.
- Morimoto, S., Kawamukai, H., and Shin, K. (2019). Prediction of crime occurrence using information propagation model and gaussian process. pages 80–87.
- Peng, C., Xueming, S., Hongyong, Y., and Dengsheng, L. (2011). Assesing temporal and weather influences on property crime in beijing, china. *Crime Law Soc Change*, 55:1–13.
- Reinhart, A. (2018). *Point Process Modeling with Spatiotemporal Covariates for Predicting Crime*. PhD thesis, Carnegie Mellon University.
- Rountree, P. W. and Land, K. C. (2000). The generalizability of multilevel models of burglary victimization: A cross-city comparison. *Social Science Research*, 29:284–305.
- Shirota, S. and Gelfand, A. E. (2017). Space and circular time log gaussian cox processes with application to crime event data. *The Annals of Applied Statistics*, 11(2):481–503.
- Short, M., D'Orsogna, M., Brantingham, P., and Tita, G. (2009). Measuring and modeling repeat and near-repeat burglary effects. *Journal of Quantitative Criminology*, 25:325–339.
- Solomon, A., Kertis, M., Shapira, B., and Rokach, L. (2022). A deep learning framework for predicting burglaries based on multiple contextual factors. *Expert Systems with Applications*, 199:117042.
- Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):111–147.

- Sypion-Dutkowska, N. and Leitner, M. (2017). Land use influencing the spatial distribution of urban crime: A case study of szczecin, poland. *ISPRS International Journal of Geo-Information*, 6(3).
- Taddy, M. A. (2010). Autoregressive mixture models for dynamic spatial poisson processes: Application to tracking intensity of violent crime. *Journal of the American Statistical Association*, 105:492:1403–1417.
- Townsley, M. K., Homel, R., and Chaseling, J. (2003). Infectious burglaries: A test of the near repeat hypothesis. *British Journal of Criminology*, 43.
- Wang, D., Ding, W., Lo, H., Stepinski, T., Salazar, J., and Morabito, M. (2013). Crime hotspot mapping using the crime related factors—a spatial data mining approach. *Appl Intell*, 39:772–781.
- Yang, X. (2006). *Exploring the influence of environmental features on residential burglary using spatial -temporal pattern analysis*. PhD thesis, University of Florida.
- Yue, H., Zhu, X., Ye, X., and Guo, W. (2017). The local colocation patterns of crime and land-use features in wuhan, china. *ISPRS International Journal of Geo-Information*, 6(10).
- Zhuang, J. and Mateu, J. (2019). A semiparametric spatiotemporal hawkes-type point process model with periodic background for crime data. *Journal of the Royal Statistical Society*, 182(3):919–942.

## A Features Preprocessing and Selection

Name	Dimension	Frequency	Description	Source	Availability
weekend	Boolean	Daily	Weekdays or Weekend indicator	Programmed	Yes
holiday	Boolean	Daily	Public holiday indicator	Holidays API	Yes
tavg	°C	Daily	Average temperature	Meteostat	Yes
prcp	mm	Daily	Average precipitation	Meteostat	Yes
year	Integer	Daily	Year indicator	Programmed	Yes
month	Integer	Daily	Month indicator	Programmed	Yes

Table 4: Overview of temporal features in the dataset.

Name	Dimension	Frequency	Description	Source	Availability
prior1d	Integer	Daily	Offenses in last 1 days (500m radius)	MyAbi	Yes
prior2d	Integer	Daily	Offenses in last 2 days (500m radius)	MyAbi	Yes
prior3d	Integer	Daily	Offenses in last 3 days (500m radius)	MyAbi	Yes
prior4d	Integer	Daily	Offenses in last 4 days (500m radius)	MyAbi	Yes
prior5d	Integer	Daily	Offenses in last 5 days (500m radius)	MyAbi	Yes
prior6d	Integer	Daily	Offenses in last 6 days (500m radius)	MyAbi	Yes
prior7d	Integer	Daily	Offenses in last 7 days (500m radius)	MyAbi	Yes

Table 5: Overview of crime-related features in the dataset.

Name	Dimension	Frequency	Description	Source	Availability
poi education	points/ hectare	yearly	Count - Point of Interest: Kindergarten, School, College, University, ...	OpenStreetMap Switzerland	Yes
poi entertainment	points/ hectare	yearly	Count - Point of Interest: Cinema, Theatre, Casino, Nightclub, ...	OpenStreetMap Switzerland	Yes
poi facilities	points/ hectare	yearly	Count - Point of Interest: Drinking Fountain, Bench, Toilets, ...	OpenStreetMap Switzerland	Yes
poi financial	points/ hectare	yearly	Count - Point of Interest: ATM, Bank, Bureau de Change, ...	OpenStreetMap Switzerland	Yes
poi gastronomy	points/ hectare	yearly	Count - Point of Interest: Cafe, Bar, Restaurant, ...	OpenStreetMap Switzerland	Yes
pub healthcare	boolean	yearly	Count - Point of Interest: Doctor, Veterinary, Pharmacy, Dentist, ...	Bundesamt für Wohnungssorgen	Yes
poi public service	points/ hectare	yearly	Count - Point of Interest: Prison, Police, Post Office, Fire Station, ...	OpenStreetMap Switzerland	Yes
poi shops	points/ hectare	yearly	Count - Point of Interest: Furniture, Groceries, Nails, ...	OpenStreetMap Switzerland	Yes
poi transportation	points/ hectare	yearly	Count - Point of Interest: Parking, Ferry St., Bus St., Taxi St.	OpenStreetMap Switzerland	Yes
poi waste management	points/ hectare	yearly	Count - Point of Interest: Recycling, Waste Disposal, Dump, ...	OpenStreetMap Switzerland	Yes

Table 6: Crime pattern theory features in the dataset.

Name	Dimension	Frequency	Description	Source	Availability
popdens	people / hectare	yearly	Density of total residential population	Statistik der Bevoelkerung	Yes
popbirth non CH	percent	yearly	Faction of residents not born in CH	Statistik der Bevoelkerung	Yes
popcit CH	percent	yearly	Fraction of Swiss citizens	Statistik der Bevoelkerung	Yes
popcit EU	percent	yearly	Fraction of EU citizens	Statistik der Bevoelkerung	Yes
popcit europ	percent	yearly	Fraction of other European (non-EU) citizens	Statistik der Bevoelkerung	Yes
popcit nonEur	percent	yearly	Fraction non-European citizens	Statistik der Bevoelkerung	Yes
popcit other	percent	yearly	Fraction citizens without citizenship	Statistik der Bevoelkerung	Yes
popcit divididx	real between 0 and 1	yearly	diversity of citizenship	Statistik der Bevoelkerung	Yes
pop age 1	percent	yearly	Fraction of residents 0-19	Statistik der Bevoelkerung	Yes
pop age 2	percent	yearly	Fraction of residents 20-34	Statistik der Bevoelkerung	Yes
pop age 3	percent	yearly	Fraction of residents 35-64	Statistik der Bevoelkerung	Yes
pop age 4	percent	yearly	Fraction of residents 65+	Statistik der Bevoelkerung	Yes
popage divididx	real between 0 and 1	yearly	Diversity of age	Statistik der Bevoelkerung	Yes
popmale	percent	yearly	Fraction of male residential population	Statistik der Bevoelkerung	Yes
popstab	percent	yearly	Fraction of stable residential population	Statistik der Bevoelkerung	Yes
popdens	people / hectare	yearly	Density of total residential population	Statistik der Bevoelkerung	Yes

Table 7: Social disorganisation theory (STATPOP) features in the dataset.

Name	Dimension	Frequency	Description	Source	Availability
busidens	businesses / hectare	yearly	Density of workplaces	Statistik der Unternehmensstruktur	Yes
busi sec 1	percent	yearly	Factions of business in sectors 1 - Extraction of Raw Materials	Statistik der Unternehmensstruktur	Yes
busi sec 2	percent	yearly	Factions of business in sectors 2 - Manufacturing	Statistik der Unternehmensstruktur	Yes
busi sec 3	percent	yearly	Factions of business in sectors 3 - Services	Statistik der Unternehmensstruktur	Yes
busisec dividx	real between 0 and 1	yearly	Diversity of workplaces	Statistik der Unternehmensstruktur	Yes
emplsec dividx	real between 0 and 1	yearly	Diversity of employees	Statistik der Unternehmensstruktur	Yes
empldens	employees/ hectare	yearly	Density of employees	Statistik der Unternehmensstruktur	Yes
emplmale	percent	yearly	Fraction of male employees	Statistik der Unternehmensstruktur	Yes

Table 8: Social disorganisation theory (STATEMP) features in the dataset.

Encoding	Education	Entertainment & Culture	Facilities	Financial	Gastronomy	Healthcare	Public Service	Transportation
0	2502 (92.8%)	2344 (86.9%)	2217 (82.2%)	2437 (90.4%)	2093 (77.6%)	2444 (90.7%)	2393 (88.8%)	1743 (64.7%)
1	158 (5.9%)	334 (12.4%)	207 (7.7%)	129 (4.8%)	242 (9.0%)	178 (6.6%)	266 (9.9%)	461 (17.1%)
2	27 (1.0%)	13 (0.5%)	126 (4.7%)	96 (3.6%)	181 (6.7%)	30 (1.1%)	34 (1.3%)	244 (9.1%)
3	8 (0.3%)	5 (0.2%)	89 (3.3%)	33 (1.2%)	74 (2.7%)	17 (0.6%)	-	78 (2.9%)
4	1 (0.0%)	-	9 (0.3%)	1 (0.0%)	51 (1.9%)	3 (0.1%)	1 (0.0%)	85 (3.2%)
5	-	-	6 (0.2%)	-	35 (1.3%)	-	2 (0.1%)	47 (1.7%)
6	-	-	9 (0.3%)	-	5 (0.2%)	-	-	16 (0.6%)
7	-	-	4 (0.1%)	-	15 (0.6%)	1 (0.0%)	-	11 (0.4%)
8	-	-	-	-	-	-	-	9 (0.3%)
10	-	-	10 (0.4%)	-	-	-	-	-
11	-	-	-	-	-	23 (0.9%)	-	2 (0.1%)
12	-	-	1 (0.0%)	-	-	-	-	-
13	-	-	18 (0.7%)	-	-	-	-	-

Table 9: Summary Statistics of Encodings for Burglary (2012-2021)

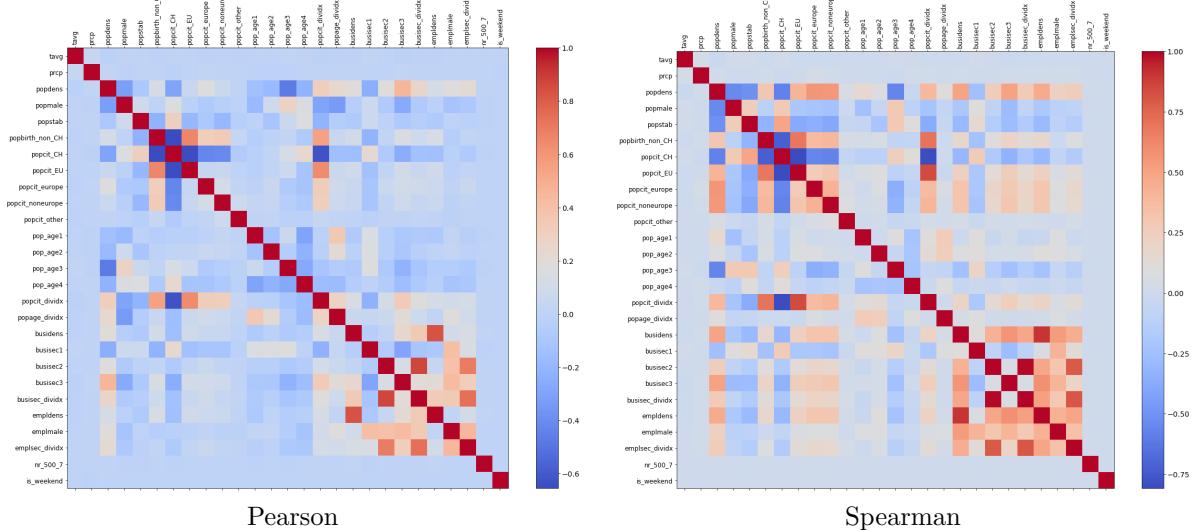


Figure 5: Feature selection: Correlation Matrices.

## B Burglary Heatmaps

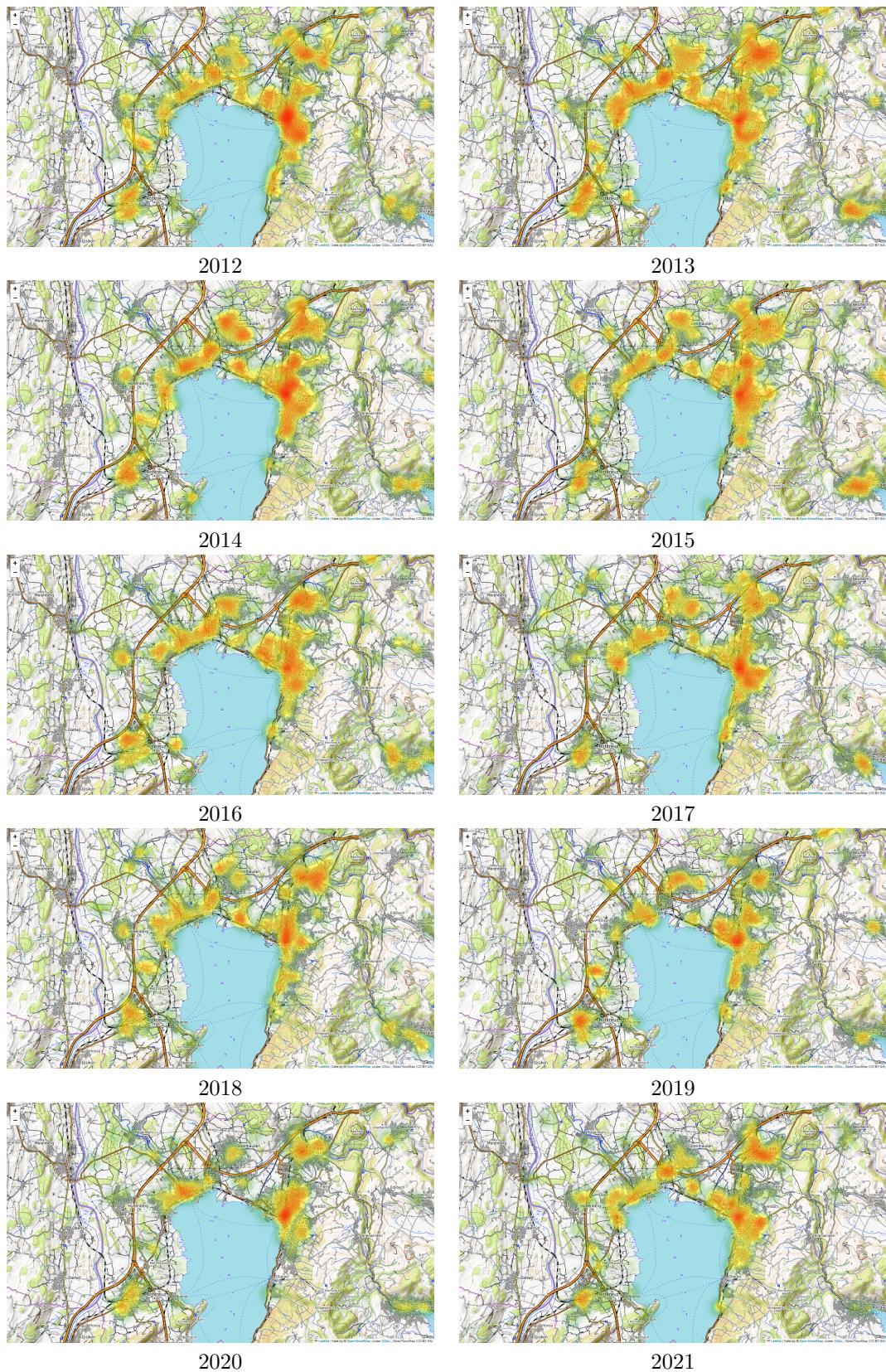


Figure 6: Heatmaps of burglaries for the years 2012 to 2021 in Canton Zug.

## C Results

Threshold	LR-L1		LR-L2		RF		AB		Ensemble	
	HR	FPR	HR	FPR	HR	FPR	HR	FPR	HR	FPR
0.10	98.10	89.45	97.93	88.28	99.65	96.76	100.00	100.00	99.14	93.65
0.15	97.41	82.62	97.06	80.17	98.27	80.61	100.00	100.00	97.06	79.55
0.19	96.20	70.66	95.34	67.30	96.89	69.06	100.00	100.00	95.85	63.79
0.24	93.61	60.12	92.75	57.27	94.82	63.14	100.00	100.00	92.23	54.82
0.29	89.64	51.92	88.43	49.15	93.61	58.05	100.00	100.00	86.70	45.45
0.34	84.80	43.67	83.07	40.18	91.88	53.58	100.00	100.00	78.76	32.54
0.38	78.76	34.16	75.47	30.33	89.29	48.22	99.31	96.51	69.60	19.76
0.43	70.98	24.78	68.39	21.32	84.28	40.25	91.54	53.75	56.82	10.18
0.48	63.56	16.90	61.14	14.18	78.41	29.75	87.56	49.35	47.50	4.86
<b>0.50</b>	<b>60.97</b>	<b>14.63</b>	<b>58.38</b>	<b>12.24</b>	<b>75.64</b>	<b>25.85</b>	<b>66.15</b>	<b>19.48</b>	<b>44.73</b>	<b>3.69</b>
0.53	56.48	10.92	54.75	9.09	69.78	18.54	46.11	3.64	40.93	2.14
0.57	50.95	6.92	49.22	5.81	56.13	8.77	33.85	0.00	39.38	1.07
0.62	45.08	4.31	43.70	3.67	45.59	3.42	33.85	0.00	36.27	0.49
0.67	41.80	2.64	41.28	2.25	41.80	1.75	33.85	0.00	34.72	0.16
0.72	40.41	1.62	39.72	1.40	37.31	0.92	33.85	0.00	33.85	0.00
0.76	37.82	1.00	37.48	0.92	33.33	0.45	33.85	0.00	33.85	0.00
0.81	36.10	0.61	35.92	0.55	30.40	0.21	33.85	0.00	33.85	0.00
0.86	35.23	0.30	34.89	0.27	17.62	0.06	17.10	0.00	33.68	0.00
0.90	34.20	0.09	34.02	0.07	9.33	0.00	0.00	0.00	31.09	0.00
0.95	33.85	0.00	33.68	0.00	3.11	0.00	0.00	0.00	6.04	0.00
1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 10: Hit Rate (HR) and False Positive Rate (FPR) in % for tested models across different thresholds.

Threshold	LR-L1		LR-L2		RF		AB		Ensemble	
	HR	FPR								
0.10	98.10	91.41	97.75	90.40	99.65	98.52	100.00	100.00	99.48	96.87
0.15	97.41	85.27	97.41	84.51	98.45	88.10	85.32	52.71	98.10	87.78
0.19	96.37	77.99	96.37	76.17	97.06	73.46	84.46	50.64	96.37	74.65
0.24	94.65	67.90	94.13	66.14	95.51	65.92	84.46	50.64	95.16	64.66
0.29	92.06	59.86	91.36	58.23	93.96	61.05	84.46	50.64	92.92	58.52
0.34	89.12	52.71	88.26	51.19	92.40	56.79	84.46	50.64	89.98	54.20
0.38	86.36	46.21	84.46	44.68	89.64	52.67	84.46	50.64	87.91	51.22
0.43	80.83	39.21	79.62	37.46	86.53	47.70	84.46	50.64	85.49	47.37
0.48	73.75	31.84	73.06	30.11	82.04	41.35	84.46	50.64	80.66	41.26
<b>0.50</b>	<b>71.67</b>	<b>29.40</b>	<b>70.80</b>	<b>27.72</b>	<b>80.31</b>	<b>38.83</b>	<b>84.46</b>	<b>50.64</b>	<b>79.10</b>	<b>38.55</b>
0.53	67.18	24.82	65.11	23.27	76.34	33.53	84.46	50.64	74.78	33.03
0.57	56.65	18.43	54.92	17.11	68.05	24.64	84.46	50.64	68.05	24.37
0.62	47.50	13.05	45.25	12.08	54.06	15.08	84.46	50.64	56.30	16.14
0.67	39.03	8.89	38.17	8.24	33.68	6.84	84.46	50.64	41.28	9.37
0.72	30.74	5.80	28.32	5.45	23.49	2.80	84.46	50.64	30.22	5.24
0.76	23.49	3.70	22.80	3.48	16.75	1.52	84.46	50.64	22.11	2.85
0.81	18.31	2.22	18.31	2.18	11.57	0.81	81.87	47.83	16.58	1.54
0.86	15.20	1.34	15.37	1.32	8.29	0.45	81.87	47.83	12.44	0.72
0.90	10.02	0.72	10.36	0.71	2.94	0.17	0.00	0.00	6.39	0.27
0.95	3.11	0.18	4.49	0.17	0.35	0.01	0.00	0.00	0.00	0.00
1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 11: Hit Rate (HR) and False Positive Rate (FPR) in % for tested models across different thresholds without the Near Repeat effect.