3D LOCAL CRIME TYPE MODELS BASED ON CRIME HOTSPOT DETECTION

E. Uglickich,[*] *I. Nagy*[*†]

Abstract: This paper deals with the analysis of the relationship between locations and types of crime observed in the Czech Republic. Cluster analysis of crime data based on the recursive Bayesian mixture estimation algorithm is used to identify crime hotspots and estimate local models of crime type. The experiments report that the 2D configuration of the algorithm allows the detection of crime hotspots online. The 3D configuration provides 29% more accurate crime type models than 2D clustering and alternative data mining algorithms. For the data set used, it was determined in which crime hotspots the most serious and most frequent types of crime can be expected to occur with the highest probability. The limitation of the study is the artificial support of the 3D clusters by the fully continuous data vector with the recoded values of the crime type. The potential use of the algorithm is expected in online web applications for sharing information on criminal offenses managed by the Police of the Czech Republic with the public and local government entities in the Czech Republic.

Key words: *crime location, crime type, cluster analysis, recursive Bayesian mixture estimation*

Received: April 20, 2023 **DOI:** 10.14311/NNW.2024.34.006
Revised and accepted: April 24, 2024

1. Introduction

The paper deals with the analysis of crime data recorded over a period of time in the Czech Republic. The available records of criminal activity to be analyzed include locations and types of crime, and the main focus of the study is to model the relationship between them. The crime records are multimodal in nature, which means that they naturally form clusters. Cluster analysis of crime data is a powerful tool for detecting crime hotspots, comparing crime trends over time within and between local regions, and predicting crime. It plays an important role in government crime prevention programs and helps improve the quality of life for citizens. Examining the behavior of the data within detected clusters helps to assess whether types of criminal activity differ between clusters. The well-interpretable crime type model

[*]Evženie Uglickich – Corresponding author; Ivan Nagy; Department of Signal Processing, Institute of Information Theory and Automation of the Czech Academy of Sciences, Pod Vodárenskou věží 4, 18208 Prague, Czech Republic, E-mail: uglickich@utia.cas.cz, nagy@utia.cas.cz

[†]Ivan Nagy; Faculty of Transportation Sciences, Czech Technical University, Na Florenci 25, 11000 Prague, Czech Republic

is expected to help identify locations with significant levels of crime severity or frequency. Records of criminal activity are both continuous and discrete in nature, making the task of constructing a model that explains the naturally non-linear dependencies between them challenging.

In exploring the state of the art in this area, it was found that recent surveys of crime data analysis and prediction [10, 14, 50, 58] mention a significant number of studies based on data mining and machine learning techniques. These include, but are not limited to, the naive Bayes classifier [44,60]; clustering algorithms such as k-means [21,63], fuzzy c-means [51], DBSCAN [5,40] and hierarchical clustering [7]; k-nearest neighbors predictor [17, 36]; decision trees and random forests [23, 49]; support vector machines [1,35] and various hybrid approaches.

Deep learning algorithms represent a powerful extension of machine learning and have attracted increasing attention in recent years in the field of crime analysis and prediction, see, e.g., [18,29,30,43,48,61,62], etc. The popularity of machine learning, and the growing attractiveness of deep learning algorithms, is understandable given the high predictive accuracy that these methods offer in various application areas [6, 13, 37, 59]. However, the poor interpretability and the high computational cost are known to be their disadvantages [37]. In contrast, statistical approaches used for crime analysis and prediction, such as, e.g., [2, 4, 27], offer good interpretability of results along with low computational complexity at the cost of relatively lower accuracy.

This paper focuses on statistical cluster analysis of crime data. The model-based clustering tools of the recursive Bayesian mixture estimation theory [33,34,41] are applied here to the task of detecting hotspots with high levels of criminal activity. The possibility of one-pass recursive estimation, free from iterative computations and increasing computational complexity during algorithm execution, is a key feature of the methodology adopted [33]. Its specific extensions have been successfully applied in various practical fields such as fuel consumption minimization [52], driving style recognition [53, 54], car accident prediction [32], transport demand prediction [47], commodity price prediction [16], etc. This paper will use the benefits of the above theory and its extension to specific crime data to demonstrate how crime hotspots can be identified in real time using actual measured data.

The specific goals of the study are:

- construct data-based models of clusters that mark crime hotspots,
- explore the types of crime in each cluster,
- construct the model describing the relationship between location and crime types,
- and validate the model on real data.

The layout of the paper is as follows: Section 2 specifies a problem and provides the theoretical background of the proposed solution. Section 3 demonstrates the application of the described algorithms to the analysis of crime data. Conclusions can be found in Section 4.

2. Theoretical Background

2.1 Problem Formulation

Consider the following records of criminal activity, marked with the time index $t = \{1, 2, \ldots, T\}$, representing the date of the crime:

- the two-dimensional continuous column data vector $\mathbf{y}_t$, whose entries are the anonymized X and Y coordinates of the crime location,
- and the discrete variable z_t, which is the type of crime with a set of possible values $\{1, 2, \ldots, N_z\}$ sorted in ascending order according to the severity of the crime.

The behavior of the observed variables changes, switching between locations where crimes are reported with different frequencies and severities, and creating clusters of values that belong to a crime hotspot. The number of these hotspots is denoted by N.

The appropriate tool to describe the behavior of such data is a mixture of N models, known as the universal approximation [22] of the multimodal relationships between the modeled variables. Fundamental approaches existing in the field of mixture estimation are based on: (i) the iterative expectation maximization (EM) algorithm [26], (ii) variational Bayes methods [25, 38, 55], (iii) numerical iterative techniques based on Markov chain Monte Carlo methods [11, 15, 19, 20], and (iv) recursive Bayesian mixture estimation [33, 34, 41] used in this paper. Bayesian mixture models have been used to analyze crime data [24, 39, 56], but to the best of our knowledge not with the recursive mixture estimation methodology mentioned above.

According to the methodology used, the mixture model of the crime data to be analyzed consists of N *components*, each in the form of a conditional probability density function (PDF) denoted by $f(\cdot|\cdot)$. In general, the ith component $\forall i \in \{1, 2, \ldots, N\}$ is the PDF describing the behavior of the data in the ith crime hotspot

$$f(\mathbf{data}_t|\Theta, c_t = i), \tag{1}$$

where $\mathbf{data}_t$ is the vector of crime records marked by the time index t, and Θ is a collection of unknown parameters of the mixture model. c_t is a *pointer* [34] whose value indicates the active component to which the data with the current index t belong. The role of the pointer c_t in this study is to mark the actual crime hotspot, while the hotspots are modeled by the components.

With the mixture model introduced, the task verbally formulated in Section 1 is specified as follows:

- identify the components describing the crime hotspots, i.e., recursively estimate the mixture parameters using past data and real-time observations,
- determine the pointer estimate pointing to the hotspot at each time t,
- construct and estimate the crime type model for each hotspot,
- validate the model.

The theoretical background to the solution of this problem is given below.

2.2 Recursive Bayesian Mixture-based Clustering

Within the framework of the adopted approach [33, 34, 41], the posterior PDF of the unknown parameters Θ and the pointer c_t is obtained using the mixture estimation algorithm based on the following *general* scheme using the Bayes rule and decomposition via the chain rule [46]:

$$\begin{aligned}\underbrace{f(\Theta, c_t = i|\{\mathbf{data}_t\}_{t=0}^{T})}_{posterior\ PDF} &\propto \underbrace{f(\mathbf{data}_t, \Theta, c_t = i|\{\mathbf{data}_t\}_{t=0}^{T-1})}_{joint\ PDF} \\ &= \underbrace{f(\mathbf{data}_t|\Theta, c_t = i)}_{component\ (1)} \underbrace{f(\Theta|\{\mathbf{data}_t\}_{t=0}^{T-1})}_{conjugate\ prior\ PDF} \\ &\times \underbrace{f(c_t = i|\{\mathbf{data}_t\}_{t=0}^{T-1})}_{pointer\ distribution}, \end{aligned} \tag{2}$$

which is simplified here by the unparameterized pointer distribution, and where $\{\mathbf{data}_t\}_{t=0}^{T}$ denotes the entire set of criminal activity records marked by indices t from $t = 0$ to $t = T$ with $\mathbf{data}_0$ as prior knowledge.

2.2.1 Parameter Estimation with Continuous Crime Locations

Applying the scheme (2) only to the continuous vector of crime locations $\mathbf{y}_t$, i.e., for the case where $\mathbf{data}_t \equiv \mathbf{y}_t$ and $\{\mathbf{data}_t\}_{t=0}^{T} \equiv \{\mathbf{y}_t\}_{t=0}^{T}$, assuming normality of its entries in clusters, the component (1) is the two-variate static normal distribution with the mean vector θ and the covariance matrix Σ. This means that the unknown parameters of each component are only $\Theta \equiv \{\theta, \Sigma\}$. The entries of the mean vector express the centers of the components describing crime hotspots. The conjugate prior PDF used in (2) for this case is the Gauss-inverse-Wischart (GiW) distribution [34, 46, 57]. The standard recursive update of GiW statistics based on currently observed crime activity is simplified by the use of static components as follows:

$$S_t = S_{t-1} + \mathbf{y}_t, \quad \kappa_t = \kappa_{t-1} + 1, \quad R_t = R_{t-1} + \mathbf{y}_t\mathbf{y}_t', \tag{3}$$

where S_0, κ_0 and R_0 are initial statistics, $'$ denotes the transposition, and which leads to a recursive recomputation of the point estimates of the mean vector θ and the covariance matrix Σ

$$\hat{\theta}_t = \frac{S_t}{\kappa_t} \equiv \frac{\sum_{t=0}^{T} \mathbf{y}_t}{T}, \quad \hat{\Sigma}_t = \frac{R_t}{\kappa_t} - \frac{S_t}{\kappa_t}\left(\frac{S_t}{\kappa_t}\right)'. \tag{4}$$

The computations (3) and (4) are performed at each time index t for each component denoting the ith crime hotspot with $i \in \{1, 2, \ldots, N\}$. See details in [33, 34, 41, 46, 57] and derivations in Appendix A.

2.2.2 Parameter Estimation with Discrete Crime Type

In the case of applying the scheme (2) only to the type of crime z_t, i.e., with $\mathbf{data}_t \equiv z_t$ and $\{\mathbf{data}_t\}_{t=0}^{T} \equiv \{z_t\}_{t=0}^{T}$, the component (1) is the $(1 \times N_z)$-dimensional categorical distribution of probabilities $\beta = \{\beta_j\}_{j=1}^{N_z} \equiv \Theta$ corresponding to the type of crime $z_t = j, \forall j = \{1, 2, \ldots, N_z\}$. The conjugate prior PDF used in (2) to estimate β is the Dirichlet distribution [33]. The recursive update of the Dirichlet statistics is fulfilled according to [33] as the stepwise construction of the $(1 \times N_z)$-dimensional contingency table ν_t as follows:

$$\nu_{j;t} = \nu_{j;t-1} + 1 \text{ for } j = z_t \text{ measured at time index } t, \tag{5}$$

where $\nu_0 = [\nu_{1;0}, \ \nu_{2;0}, \ \ldots, \ \nu_{N_z;0}]$ is zero or random initial statistics. The recomputation of the point estimates of the probabilities β of crime types is obtained by normalizing the updated statistics [33]

$$\hat{\beta}_{j;t} = \frac{\nu_{j;t}}{\sum_{l=1}^{N_z} \nu_{l;t}}. \tag{6}$$

Similar to the previous section, the recursive computations (5) and (6) are performed for each component denoting the ith crime hotspot with $i \in \{1, 2, \ldots, N\}$.

2.2.3 Active Crime Hotspot Estimation

The above formulas (3) and (4) for the crime locations or (5) and (6) for the crime type are valid for the recursive estimation of the mixture parameters, if it is known from which hotspot the crime records originate. In reality, the currently active crime hotspot has to be estimated. Here it is determined by the point estimate of the pointer c_t.

As mentioned above, the scheme (2) includes the unparameterized pointer distribution $f(c_t = i|\{\mathbf{data}_t\}_{t=0}^{T-1})$. It is the weighting vector $w_t = [w_{1;t}, \ \ldots, \ w_{N;t}]'$ obtained via the *proximity* function [31, 41, 42], which gives the approximate distance between the current crime record and each hotspot expressed by the component. The $(1 \times N)$-dimensional proximity vector m_t is obtained at each time t by substituting the parameter point estimates available from the previous time index $t-1$ and the currently measured data into each ith component (1)

$$m_{i;t} = f(\mathbf{data}_t|\hat{\Theta}_{t-1}, c_t = i), \ i \in \{1, 2, \ldots, N\}, \tag{7}$$

which is either the normal distribution with $\mathbf{y}_t$ and point estimates $\hat{\theta}_{t-1}$ and $\hat{\Sigma}_{t-1}$ or the categorical distribution with z_t and $\hat{\beta}_{t-1}$. The resulting vector m_t is normalized to obtain the weights

$$w_{i;t} = \frac{m_{i;t}}{\sum_{l=1}^{N} m_{l;t}}, \tag{8}$$

that belong to the weighting vector and represent the probabilities of crime hotspots. The crime hotspot with the highest probability among the weights is declared active at time index t, meaning that current records of crime activity belong to this hotspot. Formally, this is provided by the point estimate of the pointer c_t [34]

$$\hat{c}_t = \arg \max_{i \in \{1, \ldots, N\}} w_t. \tag{9}$$

The resulting weights (8) are used in the updates (3) and (5) of the ith component to multiply the data measured at time index t. In this way, general key steps of recognition of active crime hotspot using the given calculations are:

1. Initialize the number of components N and the initial statistics S_0, κ_0, R_0 or ν_0 for each component.

2. Compute the initial point estimates of all parameters of all components using (4) or (6).

3. **for** $t = 1, 2, \ldots$

   (a) Measure the actual $\mathbf{data}_t$.

   (b) Compute the proximity vector m_t via (7).

   (c) Compute the weighting vector w_t using (8).

   (d) Determine the point estimate of the pointer c_t via (9) to classify the data to the active crime hotspot.

   (e) Update the statistics with the help of (3) and (5) using the weighted data.

   (f) Recompute the point estimates of all parameters of all components using (4) and (6).

   (g) Go to step (a).

Note that the time loop formally runs until $t = T$, depending on the size of the data set used. However, due to the one-pass nature of the algorithm, the time t is not limited, allowing the algorithm to be applied online in real time as the actual crime records are received and analyzed. To demonstrate this, the algorithm was implemented in the free and open source programming environment Scilab (www.scilab.org) and the programming language Python 3 (www.python.org). In the next section, various configurations of the algorithm are compared and the results of crime hotspot detection are presented.

3. Application to Crime Data Analysis

3.1 Data

The records from the public database provided by the crime map web application at https://kriminalita.policie.cz are used under the terms of the general license stated on the web page. The application was created by the Police of the Czech Republic for the purpose of sharing information about the location and type of criminal activity with local government entities. The records are regularly updated on a monthly basis. This paper uses the actual data set as of August 2022. The raw data set contained 57.630 records of criminal activities observed on the territory of the Czech Republic for the month, including traffic accidents and property crimes, which were sorted out. After sorting, the data set contains 6.945 records of the anonymized coordinates of the crime location $\mathbf{y}_t$ and the crime type z_t (see Section 2.1). The following seven types of crimes remained in the data set after sorting: $z_t = 1$ – general dangerous crime, $z_t = 2$ – dangerous drugs, $z_t = 3$ – extremism, $z_t = 4$ – theft, $z_t = 5$ – burglary, $z_t = 6$ – weapons, $z_t = 7$ – violent crime.

3.2 Results

The mixture initialization was performed using the k-means clustering algorithm [28, 45] to find the initial centers of the crime hotspots. The number of hotspots was initialized to 14, corresponding to the number of regions in the Czech Republic. Then, the initial statistics and point estimates of the component parameters were computed using shuffled prior values from the initialized clusters. A different number of hotspots can be selected for initialization depending on multimodality of data sets to be analyzed.

The goal of the experiments was to apply the discussed algorithm to (i) online detection of crime hotspots based on actual measured data and (ii) estimation of the model of the type of crime in the detected hotspots.

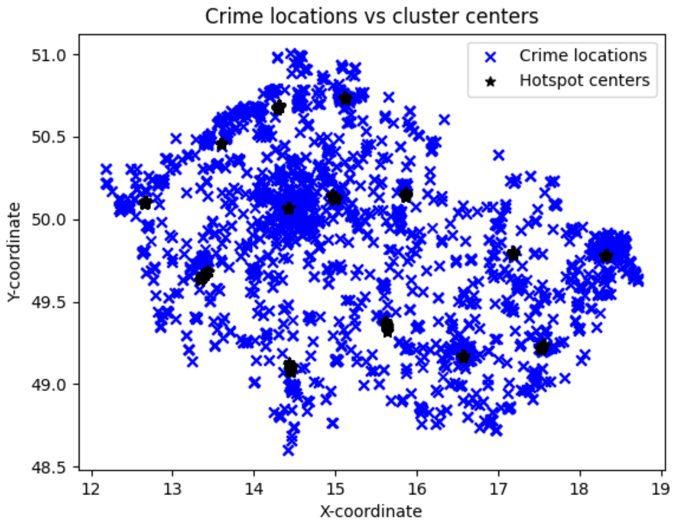
3.2.1 Online Crime Hotspot Detection

During the time loop of the one-pass algorithm, the crime hotspots were detected online using the currently measured data for each time index t. The clustered two-dimensional data space of crime locations is shown in Fig. 1.

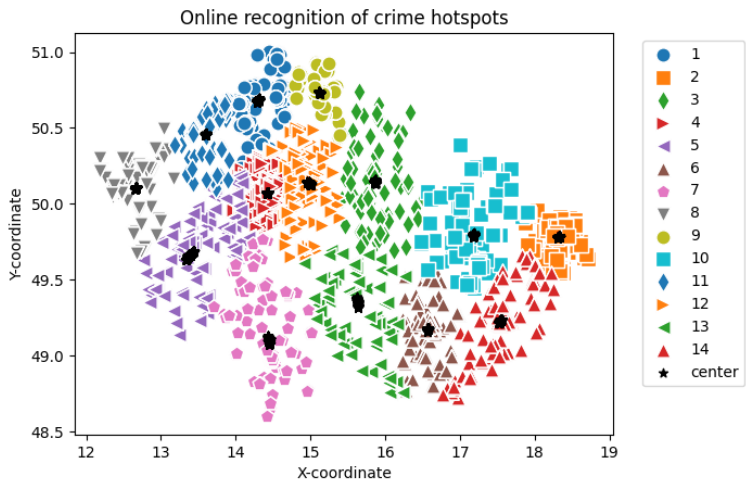
Fig. 1a shows the scatter plot of all coordinates of the crime locations in the form of a silhouette of the map of the Czech Republic. Denser clusters in the data space indicate locations with more intense criminal activity. Some of the clusters are not as dense, but the algorithm detected them as well. 14 black dots show centers of detected clusters of criminal activity. The centers are determined as the point estimates of the component parameters θ according to (4) and step (f) of the algorithm presented in Section 2.2.3. Meanwhile, the active hotspot is detected for each time index via the point estimate of the pointer c_t given by (9) and step (d) of the algorithm. Since the parameter point estimates represent the mean values of the crime coordinates during online clustering, they naturally exhibit deviations at each time index. This explains the presence of multiple black dots at each center in Fig. 1a.

In Fig. 1b, these clusters are marked with different colors along with the centers. They express the detected actual crime hotspots. Note that the boundaries of the hotspots do not necessarily correspond to the regions of the Czech Republic from which they were initialized, and some of them (4, 5 and 12) overlap. This is a natural result of the calculation of hotspot weights. An important remark is that the clusters captured are insignificantly different from those detected by centroid-based clustering methods (such as k-means used in the crime map web application at https://kriminalita.policie.cz), but the difference lies in the one-pass approach to estimating hotspot centers in the online mode. The algorithm is used to identify the actual crime hotspots in real time, providing information about the cluster to which the current crime record belongs, despite its geographic (here anonymized) coordinates.

During online clustering, all components expressing crime hotspots should be represented to verify that the model is correctly initialized. Fig. 2 shows a fragment of the evolution of 500 values of weights of hotspots at the beginning of the time loop of the algorithm execution. Note that the weights of all hotspots show their activity, i.e., the model choice is adequate, and the initialization was successful.



(a) Centers of 14 crime hotspots.



(b) 14 recognized crime hotspots.

Fig. 1 *Crime hotspots detected online. Note that hotspots create locations in the shape of a Czech Republic map silhouette.*

3.2.2 3D Local Crime Type Models Based on Crime Hotspot Detection

The correlation between crime location and crime type is low: the Spearman's rank correlation coefficient and p-value between X coordinates and crime type z_t are 0.238 and 0.0471, respectively, and they are -0.0413 and 5.7296e-04 for Y coordinates and crime type, respectively. The use of weak dependence between the variables directly for constructing a model will lead to inaccurate model and

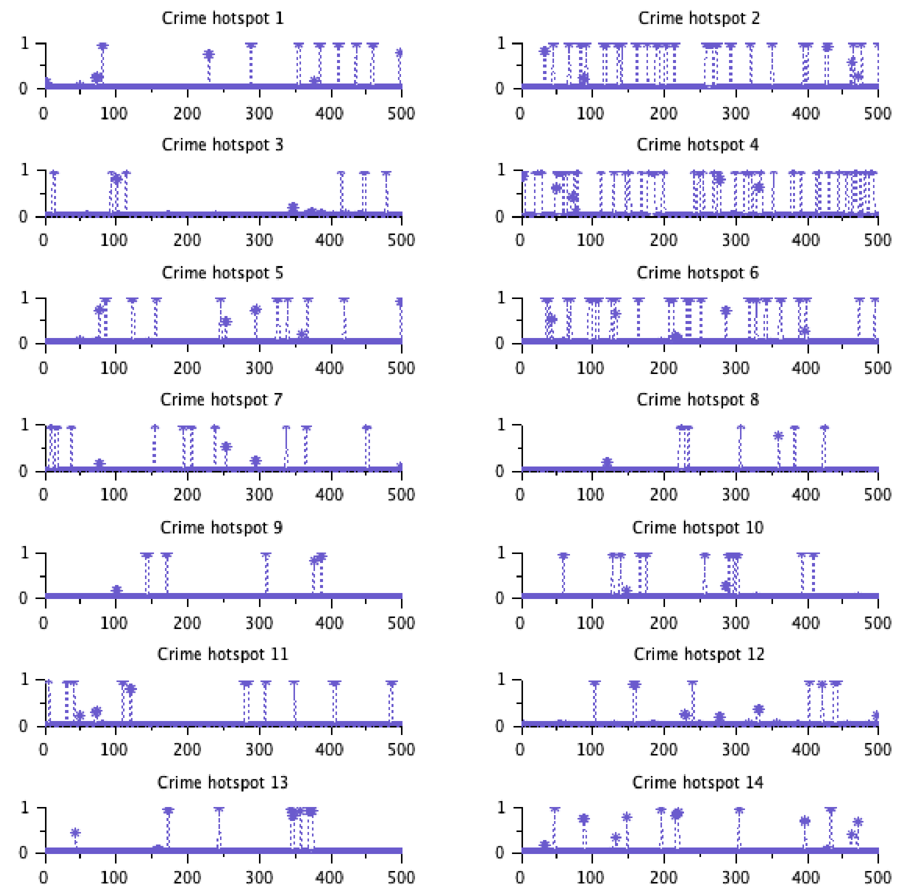
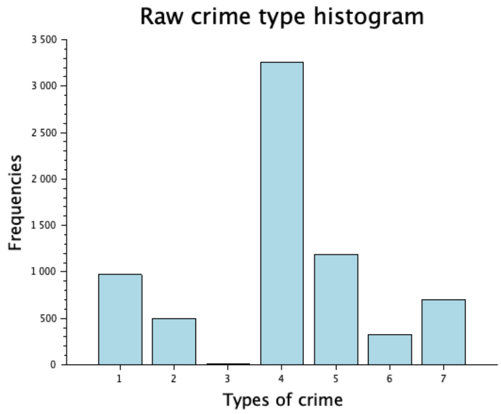


Fig. 2 *A fragment of 500 time indices (x-axis) of the online evolution of crime hotspot weights. Note that the weights of all 14 crime hotspots take on values between 0 and 1 (y-axis) as the algorithm runs. This indicates that all of the detected hotspots are regularly active.*

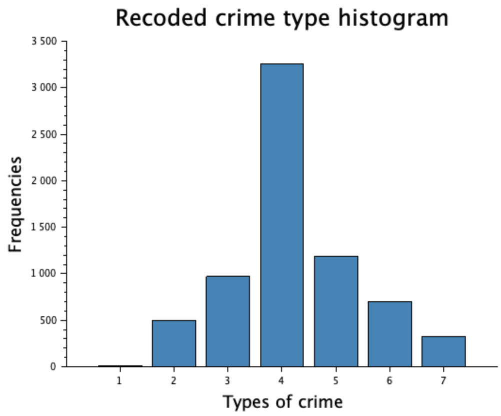
low accuracy in predicting the type of crime based on the locations. In order to address this problem, this section constructs local models of crime type that allow the relationship between variables to be captured in locations defined by crime hotspots. The individual components describing crime hotspots serve as locations of criminal activity. Using the detected locations, the local models of the type of crime have been estimated.

In this section, in order to construct the local models, the values of the crime type z_t from the data set were first reordered and recoded in the following way: $1 \rightarrow 3, 3 \rightarrow 1, 6 \rightarrow 7, 7 \rightarrow 6$. The reason for the recoding is as follows. The comparison of the raw and reordered values is given in Fig. 3. Note that while the histogram shape of the raw data in Fig. 3a corresponds to a categorical probability distribution of 7 crime types, the histogram of the recoded crime type values in Fig. 3b has a normal distribution shape due to the auxiliary reordering of the values.

Simply using the categorical distribution from Fig. 3a to predict the type of crime by maximum probability would only predict the value $z_t = 4$, which is theft, and never $z_t = 3$ (extremism). However, this auxiliary rearrangement gives us a way to deal with the recoded crime type values as continuous during computations.



(a) Histogram of raw crime type values with categorical distribution shape

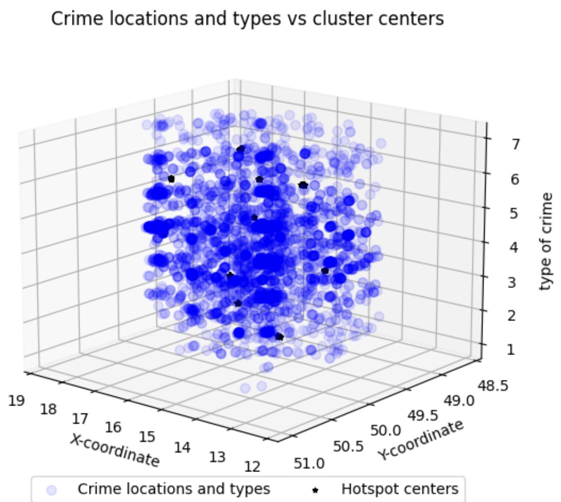


(b) Histogram of recoded crime type values with normal distribution shape.

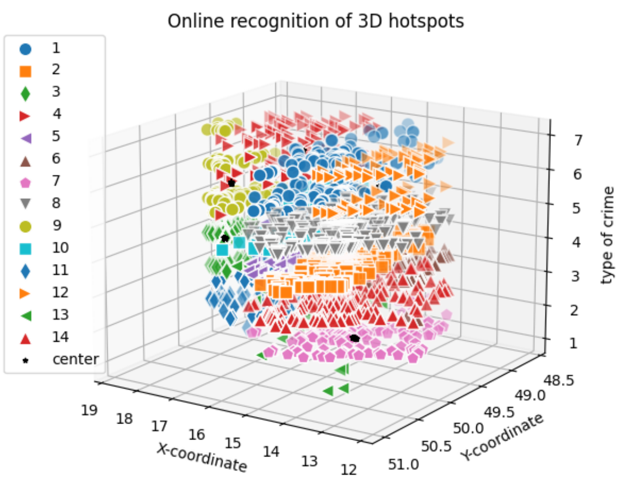
Fig. 3 *Comparison of histograms of crime type before and after recoding. Note that the values 1 and 3, and 6 and 7 have been swapped.*

This trivial recoding allows us to find 3D clusters of criminal activity in the continuous data space. Here, this is done by treating the fully continuous data vector $\mathbf{data}_t$, which contains both the crime coordinates and the recoded crime type z_t. The recoding is used only for computational reasons to obtain the proximity vector (see formula (7) in Section 2.2.3) from the three-variate normal components, but later the values of the crime type are decoded back to the raw data. Experiments have shown that using 3D clusters of crime hotspots yields more accurate local models of crime type. This is demonstrated below.

For model validation, during each run of the algorithm's time loop, the data set was randomly split into training (80%) and testing (20%) data containing 5.556 and 1.389 shuffled crime records, respectively. The 3D clustering of the fully continuous crime data space obtained with the training data is shown in Fig. 4. Fig. 4a shows the 3D space of crime locations and types, with the centers of 14 crime hotspots marked, some of which overlap or are close to each other in the data



(a) 3D crime data space with hotspot centers. Compare this data space with its 2D version shown in Fig. 1a.



(b) 14 recognized 3D crime hotspots. Note that the layers of the 3D clusters now correspond to the type of crime. It can be seen, for example, that the clusters around the value 4 (theft) are denser than the clusters corresponding to the recoded value 1 of extremism.

Fig. 4 *3D clustering of fully continuous crime data space.*

space considered. The centers detected are different from those in Fig. 1a, even though it is the same data space, but extended by crime type. Visually bounded 3D crime hotspots with layers corresponding to individual crime types, depending on the frequency of crime types in the hotspots, are shown in Fig. 4b. It can be seen in this figure that the layers of the 3D clusters correspond to the type of crime. This explains, for example, why the clusters around the value 4 (theft) are denser than the clusters corresponding to the recoded value 1 of extremism: there are significantly higher numbers of locations in the dataset where theft was recorded than for extremism.

The obtained 3D clusters are used as new 3D crime hotspots on which the local categorical crime type models are constructed and estimated according to Section 2.2.2. Each local crime type model is the 7-dimensional probability function that exists for each of the 14 hotspots and is estimated using training data. The significant contribution of the proposed approach is clearly seen in the comparison with the local models estimated on trivial 2D crime hotspots detected in the previous Section 3.2.1.

The comparison is shown in the form of histograms in Fig. 5. Fig. 5a displays the local histograms of crime type obtained on 2D crime hotspots from Fig. 1b. It is not difficult to see that, despite insignificant differences, the local crime type models in all these hotspots in Fig. 5a have a maximum probability of $z_t = 4$ (theft) and a negligible probability of $z_t = 3$ (extremism). This result is similar to the categorical distribution of crime type shown in Fig. 3a, and suggests that prediction with these local crime type models will lead to a dominance of 4 (theft) and no chance of predicting 3 (extremism).

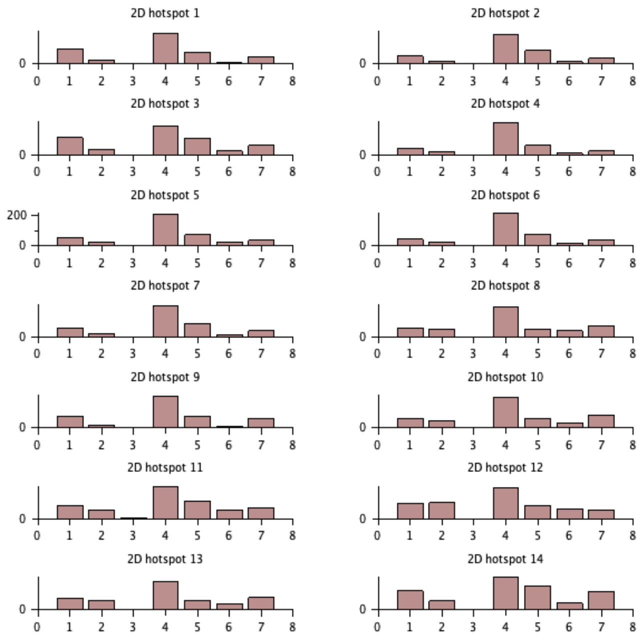
However, the 3D local crime type models shown in Fig. 5b give a completely different result. Each 3D crime hotspot provides from one to three dominant crime types with the highest probabilities differing across the hotspots, while the remaining types have negligible probabilities. To validate whether the 3D local models from Fig. 5b provide a more accurate description of the type of crime at the locations under consideration, the prediction of crime type obtained on the detected 2D and 3D crime hotspots was compared with the testing data.

Tab. I shows the comparison of the prediction accuracy defined as

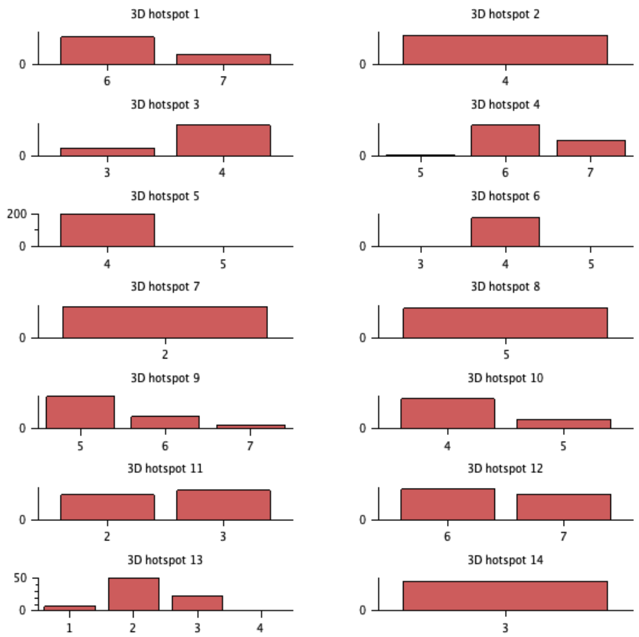
$$A = \frac{\text{number of correct predictions}}{\text{total number of predictions}} \times 100\% \tag{10}$$

averaged across 10 random splits of training and testing data. The table also provides the average prediction accuracy obtained using alternative data mining algorithms implemented in Python and KNIME (www.knime.com). The methods used for the comparison were naive Bayes (NB) [8], k-nearest neighbors (kNN) [3, 9, 12], decision trees (DT) [9, 45], random forests (RF) [9, 45], support vector machines (SVM) [9], and neural networks (NN) [9, 45]. In Tab. I, $\bar{A}$ represents the average accuracy, while σ_A indicates its standard deviation. As can be seen in Tab. I, the local models based on 3D crime hotspots show significant improvements in prediction accuracy compared to 2D and other methods. This validates the approach discussed and confirms the functionality of the crime type models.

In addition, a fragment of the visual validation of the 3D local models is shown in Fig. 6, where the evolution of the real values is compared with the obtained



(a) Local histograms of crime type on 14 2D hotspots. Note that the local models in all hotspots have a maximum probability of 4 (theft) and a negligible probability of 3 (extremism). The other crime types have insignificantly different probabilities between hotspots.



(b) Local histograms of crime type on 14 3D hotspots. Note that compared to Fig. 5a, all of these local models provide the highest probabilities of different crime types. Their number is also different. The probabilities of the remaining types are negligible.

Fig. 5 *Comparison of local crime type models estimated on 2D and 3D hotspots.*

	2D	3D	NB	kNN	DT	RF	SVM	NN
$\bar{A}(\%)$	47.230	**88.216**	46.731	58.445	52.599	52.117	19.928	47.098
$\sigma_A(\%)$	0.104	0.795	1.097	1.121	1.098	1.172	9.814	1.158

Tab. I *Average crime type prediction accuracy comparison*

predictions of the crime type. The values after decoding back to raw data are used. Due to the use of the fully continuous data vector, the crime type predictions can also be computed from the normal components on the active 3D hotspots, as shown in Fig. 6. Both categorical and continuous predictions follow the real values in the figure. The root mean square error (RMSE) of the continuous predictions averaged over 10 random splits was 0.307 with a standard deviation of 0.0099. The predictions obtained using 2D hotspots and other methods compared are concentrated around the value of 4 (theft), so there is no point in visualizing them.

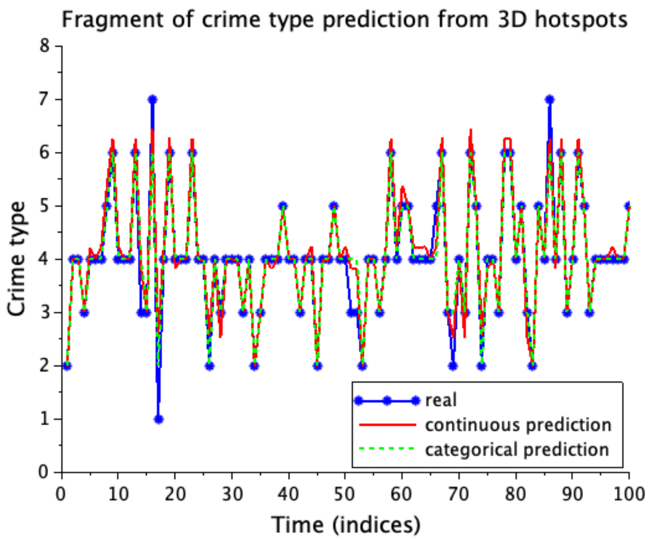
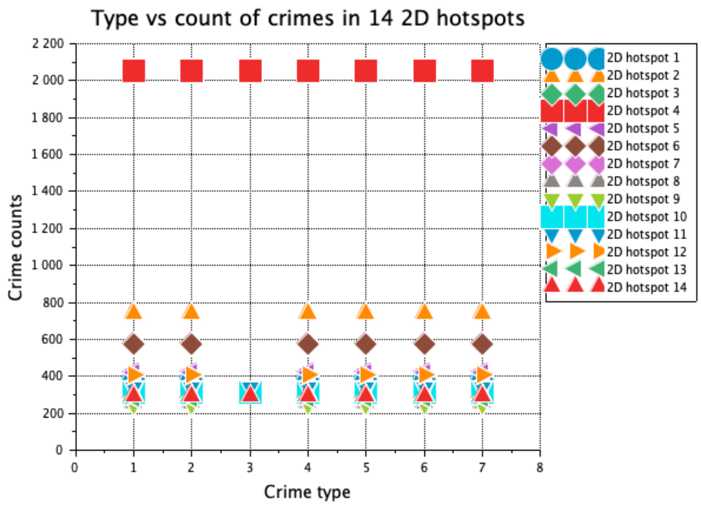
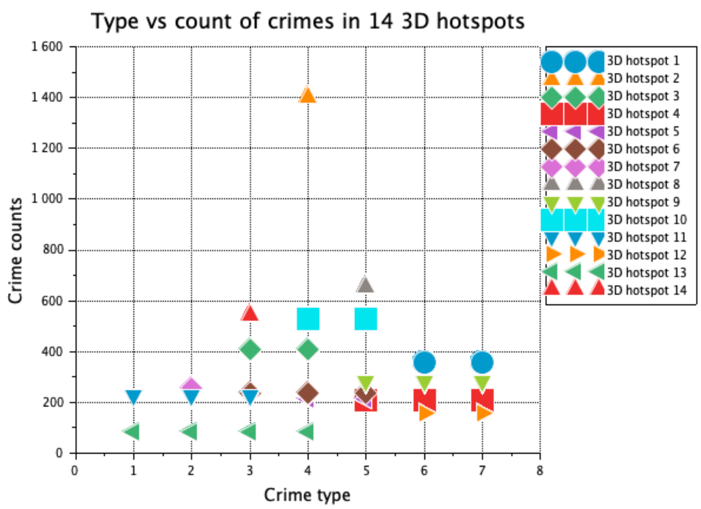


Fig. 6 *Visual comparison of real crime types and their predictions based on 3D hotspots. For better visualization, a fragment of 100 time indices from a random training set is shown. Note that both categorical and continuous predictions correspond to real values.*

The validated 3D local models describe the relationship between locations and crime types, which is the main focus of the study. The main potential of the models is seen in the search for hotspots with the most serious and/or most frequent types of crime. Fig. 7 compares the search for such crime hotspots between 2D and 3D clustering. Fig. 7a plots the type of crime against the number of all crimes in trivial 2D locations. It allows us to see that most crimes of all types are located in hotspot 4, which is around Prague in Fig. 1b, which has the shape of a map of the



(a) Note that from this plot it can be concluded that the highest counts of all crime types are in hotspot 4 around Prague (see legend of Fig. 1b). The second highest counts of all types except 3 (extremism) are observed in hotspot 2 in the Moravian-Silesian region. The two hotspots 2 and 4 are also the densest in Fig. 1a.



(b) In contrast, this plot shows that the most frequent types of crime are 4 (theft) in 3D hotspot 2 in the Moravian-Silesian region and 5 (burglary) in 3D hotspot 8 in the Karlovy Vary region. The most serious crime type 7 (violent) is reported more frequently, in descending order, in hotspot 1 in the Ústí nad Labem region, hotspot 9 in the Liberec region, hotspot 4 in Prague, and hotspot 12 in the Central Bohemia region.

Fig. 7 *Comparison of crime types and counts in 2D and 3D crime hotspots. Note that the color of each hotspot is the same as in Fig. 1b, but the shapes have been changed for better visibility.*

Czech Republic (see the legend of Fig. 1b). However, the counts of all crimes are similarly distributed across the types and hotspots, which also corresponds to the histograms in Fig. 5a.

In contrast, Fig. 7b shows specific counts of crime types in each of the 3D hotspots, allowing us to see the hotspots of the most serious crime type 7 (violent). The detailed description is given below Fig. 7b.

3.3 Discussion

The main goal of the study was to find a model that describes the relationship between crime locations and the type of crime reported in those locations. Given the weak correlation between the variables, this task was not trivial. However, it can be stated that the specific goals of the work defined in Section 1 were successfully achieved using online 3D mixture-based clustering of the crime data space. The experimental part of the study reports that the identified 3D local models accurately describe the behavior of the crime type in the detected crime hotspots in comparison to 2D local models and selected alternative well-known algorithms. The compared methods were not able to extract knowledge about the relationship between location and crime type from the raw data or from the detected 2D clusters representing crime hotspots, which explains the low prediction accuracy in validation experiments. However, the use of 3D clustering, supported by the auxiliary reordering of the crime type data, allowed the local dependencies of the variables to be captured.

Another important contribution of this approach is the online detection of crime hotspots, which applies to both 2D and 3D configurations. In fact, the application of centroid-based clustering algorithms such as k-means used for mixture initialization yields clusters insignificantly different from the discussed algorithm. However, together with the compared data mining methods, they focus more on classification problems and cross-sectional data. In contrast, the applied recursive mixture estimation is aimed more at time-series data by computing the proximity for real-time clustering and updating the weights of the distributions representing the crime hotspots.

In addition, the analysis of crime type data in 3D hotspots made it possible to search for crime hotspots with the most serious/frequent crimes, a feature that can be tailored to specific needs during data analysis.

The weakness of the approach found in the experimental part is the artificial support of 3D clusters by the fully continuous data vector with the recoded values of the crime type. This has both advantages and disadvantages. The positive side is the possibility of updating and analyzing the crime hotspot online in real time, choosing the time indices as needed. However, the online prediction of the type of crime cannot be used with the models and data considered and serves only to validate the models. A different composition of the data set will probably mitigate this shortcoming. Another possibility is to explore a dynamic model of the type of crime, which will be the subject of publication elsewhere.

The potential use of the algorithm is expected in online web applications for sharing information on criminal offenses managed by the Police of the Czech Republic with the public and local government entities in the Czech Republic, exactly in line with the objective of the original application at https://kriminalita.policie.cz, which inspired the creation of this study. Due to the contributions of the algorithm discussed above, the shared content can be enriched by the analysis of individual crime hotspots detected online.

Another direction of potential use involves monitoring the development of the crime index in hotspots of criminal activity, such as, e.g., the web application www.mapakriminality.cz, whose aim is to facilitate public orientation in the crime data regularly published by the Police of the Czech Republic.

Due to the specificity of the data used, Czech applications are mentioned. However, it should be emphasized that the data-driven approach used is not limited by data or application domain, as mentioned in Section 1.

4. Conclusion

The study dealt with a model describing the relationship between locations and types of crime in a data set observed in the Czech Republic. For the modeling, the recursive Bayesian mixture estimation algorithms were used, which have found application in many fields. The experiments conducted reported that 2D and 3D recursive clustering can be used to estimate local models of crime type corresponding to hotspots of criminal activity detected in real time, with the resulting probabilities of crime type from 2D and 3D clusters being different. The trivial 2D approach is suitable for online detection of actual crime hotspots, while the contribution of 3D local models lies in the ability to analyze the behavior of crime type in hotspots. For the data set used, it has been determined in which crime hotspots the most serious and most frequent types of crime can be expected to occur with the highest probability.

To summarize the main contributions of the proposed approach, they are: (i) online detection of crime hotspots in both 2D and 3D configurations, (ii) more accurate local models of crime type compared to alternative algorithms, and (iii) detection of hotspots with the most serious/frequent crimes.

Among the open problems that remain uncovered in the study, it is worth mentioning (i) the dynamic prediction of the type of crime, which may find a potential solution using dynamic mixtures of categorical distributions focusing on time series data, (ii) modeling the time evolution of crime counts in hotspots, which may be solved using local Poisson regressions, and (iii) predicting the status of criminal investigations. Addressing these challenges within the adopted theory of recursive Bayesian mixture estimation, which allows real-time analysis of actual measured data, will hopefully make a positive contribution to intelligent information systems to assist law enforcement agencies in identifying patterns and trends that can help prevent and analyze crimes.

Acknowledgements

European Funding: Under grant ECSEL 101007321, StorAIge is co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Chips Joint

Undertaking. Neither the European Union nor the granting authority can be held responsible for them. The project is supported by the CHIPS JU and its members (including top-up funding by France, Belgium, Czech Republic, Germany, Italy, Sweden, Switzerland, Turkey).

National Funding: ÚTIA is partner in project StorAIge. ÚTIA has received national funding from the Ministry of Education, Youth and Sports of the Czech Republic (MEYS) under grant agreement MSMT 8A21009.

The authors thank Olga Cinková for her brilliant help and fruitful discussions during the preparation of the manuscript.

Appendix A

According to [33,34,41,46,57], the relations (3) and (4) can be derived as follows. With a multivariate normal static regression model as a single component, the regression vector has the form

$$\psi = \begin{bmatrix} \mathbf{y}_t \\ 1 \end{bmatrix}, \text{ and the data matrix is } \psi\psi' = \begin{bmatrix} \mathbf{y}_t\mathbf{y}_t' & \mathbf{y}_t \\ \mathbf{y}_t' & 1 \end{bmatrix}. \tag{11}$$

The updated information matrix takes the form

$$V = \begin{bmatrix} \sum_{t=1}^{T} \mathbf{y}_t\mathbf{y}_t' & \sum_{t=1}^{T} \mathbf{y}_t \\ \sum_{t=1}^{T} \mathbf{y}_t' & T \end{bmatrix}, \tag{12}$$

where T is the number of data for which the update is running. The partition of the information matrix V is thus

$$V_y = \sum_{t=1}^{T} \mathbf{y}_t\mathbf{y}_t', \;\; V_{yp} = \sum_{t=1}^{T} \mathbf{y}_t', \;\; V_p = T, \;\; V_p^{-1} = \frac{1}{T}. \tag{13}$$

The recomputation of the point estimate of the mean vector θ is obtained as

$$\hat{\theta}_t = V_p^{-1} V_{yp} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{y}_t' = \bar{\mathbf{y}}_t{}', \tag{14}$$

and the point estimate of the covariance matrix Σ is recomputed as

$$\hat{\Sigma}_t = \frac{V_y - V_{yp}' V_p^{-1} V_{yp}}{T} = \frac{\sum_{t=1}^{T} \mathbf{y}_t\mathbf{y}_t' - \frac{\sum_{t=1}^{T} \mathbf{y}_t \sum_{t=1}^{T} \mathbf{y}_t'}{T}}{T}$$

$$= \frac{\sum_{t=1}^{T} \mathbf{y}_t\mathbf{y}_t'}{T} - \frac{\sum_{t=1}^{T} \mathbf{y}_t \sum_{t=1}^{T} \mathbf{y}_t'}{T^2} = \overline{\mathbf{y}_t\mathbf{y}_t'} - \bar{\mathbf{y}}_t\bar{\mathbf{y}}_t{}'. \tag{15}$$

References

[1] ADESOLA F., AZETA A., MISRA S., ONI A., AHUJA R., OMOLOLA A. Analysis of Violent Crime Dataset Using Support Vector Machine Model. In Emerging Technologies for Computing, Communication and Smart Cities. Springer, Singapore, 2022, pp. 193-201.

[2] AGARWAL S., YADAV L., THAKUR M.K. Crime Prediction Based on Statistical Models. In 2018 Eleventh International Conference on Contemporary Computing (IC3), 2018, pp. 1–3, doi: 10.1109/IC3.2018.8530548.

[3] ALTMAN N.S. An introduction to kernel and nearest-neighbor nonparametric regression, The American Statistician, 1992, 46(3), pp. 175–185.

[4] ALVES L.G.A., RIBEIRO H.V., RODRIGUES F.A. Crime prediction through urban metrics and statistical learning, Physica A: Statistical Mechanics and its Applications, 2018, 505, pp. 435–443.

[5] ANITHA A. Prediction of crime rate using data clustering technique. In Soft computing for problem solving. Springer, Singapore, 2020, pp. 443–454.

[6] ARAVIND T., SURESH P. Development of an efficient deep learning system for automatic prediction of power demand based on the forecasting of power distribution. Neural Network World, 2023, 33(6), pp. 461–479, doi: 10.14311/NNW.2023.33.025.

[7] BAQIR A., UL REHMAN S., MALIK S., UL MUSTAFA F., AHMAD U. Evaluating the performance of hierarchical clustering algorithms to detect spatio-temporal crime hot-spots. In: 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), IEEE, 2020, pp. 1–5.

[8] BASHA S.M., RAJPUT D.S., POLURU R.K., BHUSHAN S.B., BASHA S.A.K. Evaluating the performance of supervised classification models: decision tree and naïve Bayes using KNIME. International Journal of Engineering & Technology, 2018, 7(4.5), pp. 248–253.

[9] BERTHOLD M.R., CEBRON N., DILL F., GABRIEL T. R., KÖTTER T., MEINL T., OHL P., SIEB C., THIEL K., WISWEDEL B. KNIME: The Konstanz Information Miner, Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007). Springer, 2007, ISBN 978-3-540-78239-1, ISSN 1431-8814.

[10] BUTT U.M., LETCHMUNAN S., HASSAN F.H., ALI M., BAQIR A., SHERAZI H.H. Spatio-Temporal Crime HotSpot Detection and Prediction: A Systematic Literature Review. IEEE Access, 2020, 8, pp. 166553-166574.

[11] CHEN R., LIU J.S. Mixture kalman filters. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2000, 62(3), pp. 493–508.

[12] COVER T.M., HART P.E. Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 1967, 13(1), pp. 21–27.

[13] CUI J., SU L., WEI R., LI G., HU H., DANG X. EEG authentication based on deep learning of triplet loss, Neural Network World, 2022, 32(5), pp. 269–283, doi: 10.14311/NNW.2022.32.016.

[14] DENG S., NING Y. A Survey on Societal Event Forecasting with Deep Learning. 2021, ArXiv, abs/2112.06345.

[15] DOUCET A., ANDRIEU C. Iterative algorithms for state estimation of jump Markov linear systems. IEEE transactions on signal processing, 2001, 49(6), pp. 1216–1227.

[16] DRACHAL K. Forecasting selected energy commodities prices with Bayesian dynamic finite mixtures. Energy Economics, 2021, 99, 105283.

[17] EFENDI R., ISNAINI Y., RAHAYU S.W., MASRI R., SAMSUDIN N.A. Modeling Public Crime Type Using Multinomial Logistic Regression and-Nearest Neighbor: Pre-and During-Pandemic COVID-19. In International Conference on Soft Computing and Data Mining. Springer, Cham., 2022, pp. 320–327.

[18] FAN Y. Criminal psychology trend prediction based on deep learning algorithm and three-dimensional convolutional neural network. Journal of Psychology in Africa, 2021, 31(3), pp. 292–297.

[19] FRÜHWIRTH-SCHNATTER S. Finite Mixture and Markov Switching Models, 2nd Edition, Springer New York, 2006.

[20] FRÜWIRTH-SCHNATTER S. Fully bayesian analysis of switching gaussian state space models, Annals of the Institute of Statistical Mathematics, 2001, 53(1), pp. 31–49.

[21] HAJELA G., CHAWLA M., RASOOL A. A clustering based hotspot identification approach for crime prediction. Procedia Computer Science, 2020, 167, pp. 1462–1470.

[22] HAYKIN S. Neural Networks: A Comprehensive Foundation. Macmillan, 1994, New York.

[23] HOSSAIN S., ABTAHEE A., KASHEM I., HOQUE M.M., SARKER I.H. Crime prediction using spatio-temporal data. In International Conference on Computing Science, Communication and Security, Springer, Singapore, 2020, pp. 277–289.

[24] GAVIRIA A. Increasing returns and the evolution of violent crime: the case of Colombia. Journal of development economics, 2000, 61(1), pp. 1–25.

[25] GHAHRAMANI Z., HINTON G.E. Variational learning for switching state-space models, Neural Computation, 2000, 12(4), pp. 831–864.

[26] GUPTA M.R., CHEN Y. Theory and Use of the EM Method. (Foundations and Trends(r) in Signal Processing). Now Publishers Inc., 2011.

[27] ISLAM K., RAZA A. Forecasting crime using ARIMA model. arXiv preprint arXiv:2003.08006, 2020.

[28] JAIN A.K. Data clustering: 50 years beyond K-means. Pattern Recognition Letters. 2010, 31(8), pp. 651–666.

[29] JIN G., WANG Q., ZHU C., FENG Y., HUANG J., ZHOU J. Addressing crime situation forecasting task with temporal graph convolutional neural network approach. In 2020 12th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA). IEEE, 2020, pp. 474–478.

[30] JING C., LV X., WANG Y., QIN M., JIN S., WU S., XU G. A deep multi-scale neural networks for crime hotspot mapping prediction. Computers, Environment and Urban Systems, 2024, 109, 102089.

[31] JOZOVÁ Š., UGLICKICH E., NAGY I. Bayesian mixture estimation without tears. In: Proceedings of the 18th International Conference on Informatics in Control, Automation and Robotics – ICINCO, 2021, pp. 641–648, ISBN 978-989-758-522-7.

[32] JOZOVÁ Š., UGLICKICH E., NAGY I., LIKHONINA R. Modeling of discrete questionnaire data with dimension reduction. Neural Network World, 2022, 32(1), pp. 15–41, doi: 10.14311/NNW.2022.32.002.

[33] KÁRNÝ M., BÖHM J., GUY T.V., JIRSA L., NAGY I., NEDOMA P., TESAŘ L. Optimized Bayesian Dynamic Advising: Theory and Algorithms. Springer, London, 2006.

[34] KÁRNÝ M., KADLEC J., SUTANTO E.L. Quasi-Bayes estimation applied to normal mixture. In: Rojíček, J., Valečková, M., Kárný, M., Warwick, K. (eds.). Preprints of the 3rd European IEEE Workshop on Computer-Intensive Methods in Control and Data Processing, Prague, CZ, 1998, pp. 77–82.

[35] KIANMEHR K., ALHAJJ R. Effectiveness of support vector machine for crime hot-spots prediction. Applied Artificial Intelligence, 2008, 22(5), pp. 433–458.

[36] KUMAR A., VERMA A., SHINDE G., SUKHDEVE Y., LAL N. Crime Prediction Using K-Nearest Neighboring Algorithm. In: 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, pp. 1–4, doi: 10.1109/ic-ETITE47903.2020.155.

[37] LIN G., LIN A., GU D. Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient. Information Sciences, 2022, 608, pp. 517–531.

[38] MCGRORY C.A., TITTERINGTON D.M. Variational Bayesian analysis for hidden Markov models, Australian & New Zealand Journal of Statistics, 2009, 51, pp. 227–244.

[39] MELNYKOV V., ZHU X. Studying crime trends in the USA over the years 2000–2012. Advances in Data Analysis and Classification, 2019, 13(1), pp. 325–341.

[40] MUNGEKAR D., JOSHI H., KANKEKAR A., NAIR P., DAS P. Crime Analysis using DBSCAN Algorithm. In: 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), IEEE, 2021, pp. 628-635.

[41] NAGY I., SUZDALEVA E. Algorithms and Programs of Dynamic Mixture Estimation. Unified Approach to Different Types of Components, SpringerBriefs in Statistics. Springer International Publishing, Heidelberg, 2017.

[42] NAGY I., SUZDALEVA E., PECHERKOVÁ P. Comparison of various definitions of proximity in mixture estimation. In: Proceedings of the 13th International Conference on Informatics in Control, Automation and Robotics, ICINCO, 2016, 1, ISBN 978-989-758-198-4, pp. 527–534.

[43] NAIR S N., GOPI E.S. Deep learning techniques for crime hotspot detection. In Optimization in Machine Learning and Applications, Springer, Singapore, 2020, pp. 13–29.

[44] NITTA G.R., RAO B.Y., SRAVANI T., RAMAKRISHIAH N., BALAANAND M. LASSO-based feature selection and naïve Bayes classifier for crime prediction and its type. Service Oriented Computing and Applications, 2019, 13(3), pp. 187–197.

[45] PEDREGOSA F., VAROQUAUX G., GRAMFORT A., MICHEL V., THIRION B., GRISEL O., BLONDEL M., PRETTENHOFER P., WEISS R., DUBOURG V., VANDERPLAS J., PASSOS A., COURNAPEAU D., BRUCHER M., PERROT M., DUCHESNAY E. Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 2011, 12, pp. 2825–2830.

[46] PETERKA V. Bayesian system identification, in Eykhoff, P. (Ed.), Trends and Progress in System Identification. Oxford, Pergamon Press, 1981, pp. 239–304.

[47] PETROUŠ M., SUZDALEVA E., NAGY I. Modeling of passenger demand using mixture of Poisson components. In: *Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, 2019, pp. 617–624.

[48] RAJA R.A., YUVARAJ N., KOUSIK N.V. Analyses on Artificial Intelligence Framework to Detect Crime Pattern. Intelligent Data Analytics for Terror Threat Prediction: Architectures, Methodologies, Techniques and Applications, 2021, pp. 119–132.

[49] RAMASUBBAREDDY S., ADITYA SAI SRINIVAS T., GOVINDA K., MANIVANNAN S.S. Crime prediction system. In Innovations in Computer Science and Engineering. Springer, Singapore, 2020, pp. 127–134.

[50] SARAVANAN P., SELVAPRABU J., ARUN RAJ L., ABDUL AZEEZ KHAN A., JAVUBAR SATHICK K. Survey on Crime Analysis and Prediction Using Data Mining and Machine Learning Techniques. In: Zhou, N., Hemamalini, S. (eds) Advances in Smart Grid Technology. Lecture Notes in Electrical Engineering, Springer, Singapore, 688, 2021.

[51] SIVANAGALEELA B., RAJESH S. Crime analysis and prediction using fuzzy c-means algorithm. In: *3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, IEEE, 2019, pp. 595–599.

[52] SUZDALEVA E., NAGY I. Data-based speed-limit-respecting eco-driving system. Transportation Research Part C: Emerging Technologies, 2014, 44, pp. 253–264.

[53] SUZDALEVA E., NAGY I. An online estimation of driving style using data-dependent pointer model. Transportation research part C: emerging technologies, 2018, 86, pp. 23–36.

[54] SUZDALEVA E., NAGY I. Two-layer pointer model of driving style depending on the driving environment. Transportation research part B: methodological, 2019, 128, pp. 254–270.

[55] ŠMÍDL V., QUINN A. The Variational Bayes Method in Signal Processing, Springer, 2005.

[56] TADDY M.A. Autoregressive mixture models for dynamic spatial Poisson processes: Application to tracking intensity of violent crime. Journal of the American Statistical Association, 2010, 105(492), pp. 1403–1417.

[57] TITTERINGTON D.M., SMITH A.F.M., MAKOV U.E. Statistical Analysis of Finite Mixture Distributions (Wiley Series in Probability and Statistics – Applied Probability and Statistics Section). Wiley, 1986.

[58] THOMAS A., SOBHANA N.V. A survey on crime analysis and prediction. Materials Today: Proceedings, 2022, doi: 10.1016/j.matpr.2022.02.170.

[59] TURK F., AKKUR E., EROGUL O. BI-RADS categories and breast lesions classification of mammographic images using artificial intelligence diagnostic models. Neural Network World, 2023, 33(6), pp. 413–432, doi: 10.14311/NNW.2023.33.023.

[60] VURAL M.S., GÖK M. Criminal prediction using Naive Bayes theory. Neural Computing and Applications, 2017, 28(9), pp. 2581–2592.

[61] WANG B., YIN P., BERTOZZ, A.L., BRANTINGHAM P.J., OSHER S.J., XIN J. Deep learning for real-time crime forecasting and its ternarization. Chinese Annals of Mathematics, Series B, 2019, 40(6), pp. 949–966.

[62] ZHANG Y., CHENG T. Graph deep learning model for network-based predictive hotspot mapping of sparse spatio-temporal events. Computers, Environment and Urban Systems, 2020, 79, 101403.

[63] ZHU Q., ZHANG F., LIU S., LI Y. An anticrime information support system design: Application of K-means-VMD-BiGRU in the city of Chicago. Information & Management, 2022, 59(5), 103247.