

Article

Not peer-reviewed version

Integrating Spatial Clustering and Machine Learning for Territorial Segmentation and Short-Term Crime Prediction in Tamaulipas, Mexico

[Elvira Rolón](#)*, [José G. Méndez](#), [Roberto Pichardo](#)

Posted Date: 28 January 2026

doi: 10.20944/preprints202601.2193.v1

Keywords: crime analysis; territorial segmentation; spatial clustering; AdaBoost; short term crime prediction; public security; Tamaulipas



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Integrating Spatial Clustering and Machine Learning for Territorial Segmentation and Short-Term Crime Prediction in Tamaulipas, Mexico

Elvira Rolón¹, José G. Méndez¹ and Roberto Pichardo¹

Universidad Autónoma de Tamaulipas, Tampico, Mexico

* Correspondence: erolon@docentes.uat.edu.mx

Abstract

Crime prediction and territorial analysis have become increasingly relevant for public security planning, particularly in regions characterized by heterogeneous spatial and temporal crime dynamics. This study proposes an integrated methodological framework that combines spatial clustering and supervised machine learning to support territorial segmentation and short-term crime occurrence prediction in the state of Tamaulipas, Mexico. The proposed approach follows the Knowledge Discovery in Databases (KDD) process and is based on official crime records analyzed at the neighborhood (colonia) level across eleven municipalities. In the first stage, a K-Means clustering algorithm is applied to identify homogeneous territorial patterns based on crime incidence and sociodemographic characteristics. In the second stage, an AdaBoost classifier is implemented to predict the occurrence of crime events using different temporal windows. Model performance is evaluated using precision, recall, F1-score, and accuracy, with particular emphasis on recall due to the operational relevance of minimizing false negatives in public security contexts. The results indicate that the combined spatial and predictive approach supports the understanding of territorial crime dynamics and provides stable predictive performance across municipalities. This integration offers a practical and replicable framework to support data-driven decision-making in public security and territorial planning, particularly in contexts with limited analytical resources.

Keywords: crime analysis; territorial segmentation; spatial clustering; AdaBoost; short-term crime prediction; public security; Tamaulipas

1. Introduction

Crime represents one of the most pressing challenges for public security and social development, particularly in urban contexts characterized by strong territorial heterogeneity and dynamic temporal patterns. In Mexico, criminal activity remains at high levels and exhibits pronounced spatial differences across regions and municipalities, complicating the implementation of uniform prevention strategies. In this context, the increasing availability of administrative crime data has created opportunities to develop evidence-based analytical approaches to support decision-making in public security.

Traditionally, crime analysis has relied on descriptive statistical methods and explanatory models focused on socioeconomic variables. While these approaches have contributed to understanding structural factors associated with criminal activity, their capacity to capture complex, dynamic, and non-linear patterns is limited. In contrast, machine learning techniques have demonstrated substantial potential for analyzing large, heterogeneous datasets, uncovering latent regularities, and generating predictive estimates in complex environments.

Recent literature identifies two dominant methodological perspectives in data-driven crime analysis. On the one hand, territorial segmentation approaches, commonly implemented through unsupervised clustering techniques, enable the spatial characterization of areas with similar crime

profiles and support territorial diagnostics. On the other hand, supervised machine learning models have been widely applied to predict the occurrence of crime events over short-term temporal horizons, emphasizing operational anticipation. However, these approaches are often developed independently, limiting a comprehensive understanding of crime dynamics.

This study proposes an analytical framework that integrates both perspectives in a complementary manner, while maintaining a clear conceptual and methodological separation between descriptive spatial analysis and temporal prediction. The methodology follows the Knowledge Discovery in Databases (KDD) process and is based on official crime records from eleven municipalities in the state of Tamaulipas, Mexico, covering the period from 2017 to 2023.

In the first stage, the K-Means clustering algorithm is applied to segment neighborhoods within each municipality based on crime incidence and socioeconomic characteristics, with the aim of identifying homogeneous territorial patterns. In the second stage, independently of the clustering process, a supervised predictive model based on AdaBoost is implemented to estimate the probability of at least one crime event occurring within 1-, 7-, and 15-day time windows, using temporal and cumulative features derived from historical records.

The main contribution of this work lies in the articulation of a dual analytical approach that combines territorial segmentation and temporal prediction as complementary components, without imposing methodological dependence between them. This design enables the generation of interpretable insights for both spatial crime diagnosis and short-term operational anticipation, offering a replicable framework applicable to territorially heterogeneous contexts with limited analytical resources. The results show consistent territorial segmentation and stable predictive performance, supporting the relevance of the proposed approach for public security planning and crime prevention strategies.

2. Related Work

Recent literature on data-driven crime analysis can be broadly grouped into two main research lines: territorial segmentation approaches based on unsupervised learning techniques and supervised predictive models aimed at anticipating crime events.

In the first line, several studies have employed clustering algorithms, particularly K-Means, to spatially characterize criminal activity and classify regions or territorial units according to crime incidence. Works such as those by [1–3] demonstrate the usefulness of unsupervised clustering for identifying spatial patterns and grouping areas with similar crime profiles. Complementarily, [4,6] apply K-Means to segment urban areas and analyze specific crime types, highlighting the descriptive value of clustering techniques for territorial diagnostics.

Other studies have proposed variations or optimizations of the K-Means algorithm to improve clustering quality. For instance, [5] introduce an optimized K-Means approach for crime analysis and prediction. In the Latin American context, [7] present an exploratory analysis supported by machine learning techniques to characterize cybercrime, illustrating the potential of these methods in regions with high territorial heterogeneity.

A second research line focuses on the use of supervised machine learning models for crime prediction. Recent systematic reviews report a growing interest in algorithms such as decision trees, neural networks, ensemble methods, and boosting techniques to anticipate crime occurrence over different temporal horizons [8,9]. These studies emphasize the ability of supervised models to handle large datasets and capture complex, non-linear relationships.

Within this line, AdaBoost has been identified as a competitive approach for crime-related classification tasks, particularly in contexts characterized by class imbalance. [10] demonstrate the performance of a Naïve Bayes–AdaBoost ensemble for classifying sexual crimes, while other studies report favorable results using machine learning techniques for crime prediction and forecasting [11–13].

Despite these advances, most existing studies tend to develop territorial segmentation or predictive modeling approaches in isolation. Clustering-based works primarily focus on spatial characterization and descriptive analysis, without explicitly addressing short-term crime anticipation.

Conversely, predictive studies often operate on aggregated spatial units, without incorporating an explicit territorial diagnostic to contextualize their results.

In this context, a methodological gap persists in the development of approaches that combine territorial segmentation and temporal crime prediction in a complementary manner, while maintaining a clear conceptual separation between both components. The present study addresses this gap by proposing a dual analytical framework that integrates unsupervised clustering and supervised machine learning within the KDD process, offering a coherent and replicable approach for crime analysis and anticipation in territorially heterogeneous contexts.

3. Study Area and Data

The study area corresponds to the state of Tamaulipas, located in northeastern Mexico, a region characterized by pronounced territorial heterogeneity in terms of urbanization, economic dynamics, and crime patterns. For analytical purposes, this study focuses on eleven municipalities within the state, which concentrate the complete and consistent availability of crime records throughout the period of analysis. These municipalities include border areas, metropolitan regions, and semi-urban zones, allowing the capture of a wide range of territorial contexts.

The spatial unit of analysis adopted in this study is the neighborhood (colonia), considered a relevant intra-urban scale for crime analysis. This level of spatial disaggregation enables the identification of fine-grained spatial variations that are often obscured when using more aggregated administrative units. The selection of this unit is driven both by the availability of official data and its operational relevance for public security planning.

Crime data were obtained from official sources, specifically from the Executive Secretariat of the State Public Security System (SESESP) through its State Information Center. The dataset documents criminal incidents occurring between 2017 and 2023 and includes information on the event date, municipality, neighborhood, crime type, and the number of recorded occurrences. In total, the dataset comprises 48,768 records corresponding to multiple categories of criminal activity.

In addition, for the purpose of territorial segmentation, socioeconomic variables at the neighborhood level were incorporated from official statistical sources. These variables provide a structural context for crime incidence and include demographic and socioeconomic indicators commonly used in territorial studies. All socioeconomic data were spatially harmonized to ensure consistency with the crime records at the neighborhood level.

It is important to note that, although both datasets share the same spatial unit, their use within the study serves distinct analytical purposes. Socioeconomic variables are employed exclusively in the territorial segmentation model based on clustering techniques, whereas the supervised predictive model is constructed solely from temporal and cumulative features derived from historical crime records. This separation preserves methodological coherence between both approaches and avoids unnecessary analytical dependencies.

Prior to analysis, the data were subjected to cleaning and standardization procedures to ensure consistency and comparability. These processes include the removal of non-assignable records, the normalization of neighborhood names, and the consolidation of temporal variables. A detailed description of these procedures is provided in the Methodology section.

Table 1 summarizes the main types of variables used in the study and their analytical role within each modeling component.

Table 1. Summary of variables used in the study

Variable type	Description	Model usage
Crime frequency	Total crime events per neighborhood	Clustering
Socioeconomic indicators	Demographic and socioeconomic attributes	Clustering
Day of week	Temporal categorical variable	Prediction
Month	Temporal categorical variable	Prediction
Season	Temporal categorical variable	Prediction
Lagged crime counts	Crimes in previous periods (7, 15, 30 days)	Prediction
Cumulative crime measures	Historical crime intensity indicators	Prediction

4. Methodology

The methodological design of this study follows the Knowledge Discovery in Databases (KDD) framework, which provides a structured and iterative process for transforming raw data into actionable knowledge. Within this framework, the proposed approach integrates two complementary analytical components: territorial segmentation through unsupervised learning and short-term crime prediction through supervised classification. The complete workflow is summarized in Figure 1.

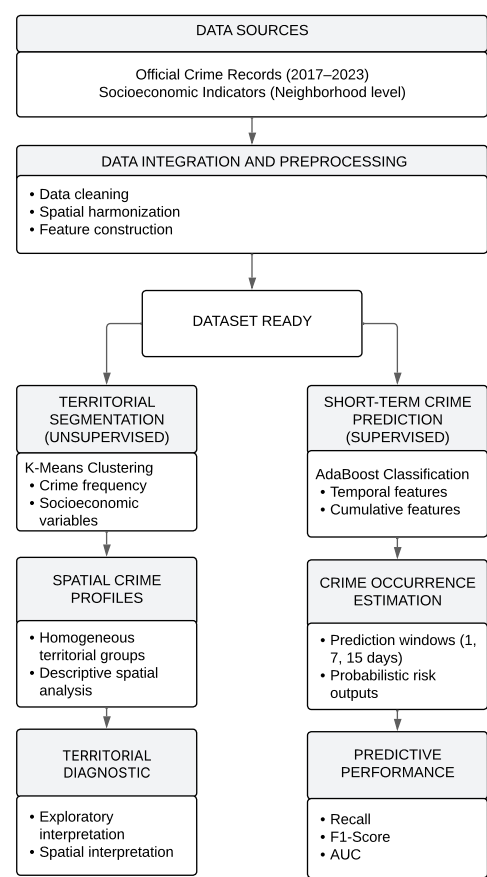


Figure 1. Methodological workflow illustrating the dual analytical approach based on territorial segmentation and supervised crime prediction.

4.1. Data Integration and Preprocessing

The analysis begins with the integration of official crime records and contextual information at the colonia level. Crime data were obtained from administrative records compiled by public security institutions, while contextual variables were derived from official statistical sources. Both datasets were spatially harmonized to ensure consistency in geographic identifiers.

Preprocessing procedures included the elimination of duplicate and non-assignable records, standardization of colonia names, and removal of special characters to ensure compatibility across sources. Records corresponding to crime categories with incomplete temporal coverage were excluded

to preserve the consistency of the analytical period. A unified date variable was constructed from year, month, and day fields to support subsequent feature engineering.

4.2. Feature Engineering

To enhance the predictive capacity of the models, a set of temporal, spatial, and cumulative features was derived from the crime records. Temporal attributes include the day of the week, month of the year, season, and week of the year. In addition, cumulative and lag-based indicators were constructed to capture recent crime dynamics within each colonia, such as the number of events observed in the previous 7, 15, and 30 days.

Contextual features include the type of urban settlement associated with each colonia and aggregate indicators reflecting the historical volume of crime. Together, these variables provide a structured representation of both short-term dynamics and longer-term crime intensity at the territorial level.

4.3. Spatial Segmentation Through Clustering

Territorial segmentation was performed using an unsupervised clustering approach to identify groups of colonias with similar crime-related characteristics. Clustering was conducted using the K-Means algorithm, selected for its efficiency and suitability for partitioning large spatial datasets. Prior to clustering, numerical variables were normalized to avoid scale effects.

Similarity between observations was measured using the Euclidean distance. Given two observations $\mathbf{x}_i$ and $\mathbf{x}_j$ described by p numerical attributes, the Euclidean distance is defined as:

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}. \tag{1}$$

The optimal number of clusters was determined using internal validation criteria, including the elbow method and the Davies–Bouldin index. The resulting clusters define a territorial typology that supports exploratory analysis and spatial interpretation. Importantly, cluster assignments were used exclusively for descriptive purposes and were not incorporated as predictors in the supervised classification models, ensuring a clear conceptual separation between segmentation and prediction.

4.4. Construction of Prospective Target Variables

Crime prediction was formulated as a binary classification problem. For each colonia and reference date, three prospective target variables were defined to indicate the occurrence or non-occurrence of at least one crime event within future time windows of 1 day, 7 days, and 15 days. This formulation enables short-term crime anticipation without relying on explicit time-series modeling, focusing instead on window-based event occurrence.

Table 2 presents the definition of the prediction windows used in the supervised classification task.

Table 2. Definition of prediction windows

Window	Prediction horizon	Target definition
Short-term	1 day	Crime occurrence within next 24 hours
Medium-term	7 days	Crime occurrence within next 7 days
Extended-term	15 days	Crime occurrence within next 15 days

Formally, for each neighborhood and reference time t , the target variable y_t is defined as:

$$y_t = \begin{cases} 1, & \text{if at least one crime event occurs in } (t, t + \Delta t], \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

where $\Delta t \in \{1, 7, 15\}$ days represents the prediction window.

4.5. Supervised Classification and Class Imbalance Handling

Supervised learning was implemented using the AdaBoost ensemble algorithm. AdaBoost iteratively combines weak learners and emphasizes misclassified observations, making it well suited for heterogeneous and imbalanced crime datasets. Separate models were trained for each temporal horizon.

To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to the training data, generating synthetic instances of the minority class and improving the model’s ability to learn patterns associated with rare events. Numerical variables were scaled using a robust normalization strategy based on the median and interquartile range, while categorical variables were encoded using one-hot encoding.

4.6. Model Training and Evaluation

Model performance was evaluated using a stratified data partitioning strategy, with 80% of the data used for training and 20% reserved for testing. Stratified k-fold cross-validation was applied during hyperparameter tuning to preserve class proportions across folds. Hyperparameters were optimized through grid search.

Evaluation metrics included accuracy, precision, recall, F1-score, and the area under the ROC curve (AUC). Given the operational relevance of crime anticipation, particular emphasis was placed on recall to minimize false negatives. Performance was assessed independently for each temporal horizon, allowing comparison of predictive behavior across short-term windows.

Given the operational relevance of minimizing undetected crime events, recall was prioritized as a key evaluation metric. Recall is defined as:

$$\text{Recall} = \frac{TP}{TP + FN}$$

(3)

where TP denotes true positives and FN false negatives.

5. Results

This section presents the results obtained from the two analytical components of the proposed framework: territorial segmentation using unsupervised clustering and short-term crime prediction using supervised machine learning. Results are reported using the best-performing configurations previously validated in independent studies and are organized according to each analytical component.

5.1. Territorial Segmentation Results

Territorial segmentation was conducted using the K-Means algorithm, enabling the identification of homogeneous neighborhood groups based on crime incidence and socioeconomic characteristics. Clustering was performed independently for each municipality to capture localized spatial dynamics and avoid imposing a uniform clustering structure across the state.

Cluster quality was evaluated using the Davies–Bouldin index, which measures intra-cluster dispersion relative to inter-cluster separation. Across the eleven municipalities analyzed, Davies–Bouldin values ranged from 0.371 to 0.856, indicating moderate to good cluster separation depending on local conditions. Lower values were observed in municipalities such as Valle Hermoso and San Fernando, suggesting more compact and well-separated clusters, whereas higher values in municipalities such as Reynosa reflect more complex or heterogeneous territorial crime patterns.

Table 3 summarizes the optimal number of clusters selected for each municipality and the corresponding Davies–Bouldin index.

Table 3. Territorial clustering results by municipality

Municipality	Optimal clusters (k)	Davies–Bouldin
Altamira	6	0.760
Ciudad Madero	5	0.768
El Mante	6	0.598
Matamoros	9	0.509
Nuevo Laredo	7	0.731
Reynosa	6	0.856
Río Bravo	7	0.769
San Fernando	6	0.508
Tampico	6	0.805
Valle Hermoso	6	0.371
Ciudad Victoria	6	0.621

The resulting clusters provide an interpretable territorial diagnostic by revealing distinct spatial patterns of criminal activity. In several municipalities, a limited number of neighborhoods concentrate a disproportionately high number of crime events, while other clusters exhibit consistently low incidence levels despite adverse socioeconomic conditions. These results highlight the usefulness of territorial segmentation as a tool for spatial crime diagnosis and localized analysis.

5.2. Predictive Performance Results

For the predictive component, supervised classifiers based on the AdaBoost algorithm were trained to estimate the occurrence of crime events within short-term temporal windows. Model performance was evaluated using unseen test data and summarized through standard classification metrics.

Overall, the predictive model exhibited stable and balanced performance. The average accuracy reached 0.77, while precision and recall values were 0.84 and 0.77, respectively. The resulting F1-score was 0.79, indicating an adequate balance between sensitivity and precision. In addition, the average area under the ROC curve (AUC) was 0.77, reflecting satisfactory discriminative capacity across municipalities and temporal windows.

Table 4 summarizes the global predictive performance of the AdaBoost model.

Table 4. Global predictive performance of the AdaBoost model

Metric	Average value
Accuracy	0.77
Precision	0.84
Recall	0.77
F1-score	0.79
AUC	0.77

The prioritization of recall contributed to reducing the proportion of false negatives, which is critical in public security contexts where failing to anticipate a crime event may have significant operational consequences. Despite variability across municipalities with different crime volumes, the model maintained consistent performance without abrupt degradation, supporting its applicability for short-term crime anticipation in territorially heterogeneous settings.

6. Discussion

The results obtained confirm the usefulness of addressing crime analysis through a dual perspective that combines territorial segmentation and temporal prediction while maintaining a clear conceptual separation between both components. Clustering-based segmentation revealed differentiated spatial patterns across municipalities, reflecting the territorial heterogeneity of crime dynamics in the state of Tamaulipas. The observed variability in Davies–Bouldin index values indicates that the

spatial structure of crime is not uniform and is strongly influenced by local conditions, supporting the relevance of intra-urban scale analysis.

From a territorial perspective, the identified clusters show that criminal activity tends to concentrate in a limited number of neighborhoods within each municipality, while extensive areas exhibit consistently low incidence levels. This pattern is consistent with findings reported in spatial criminology studies and supports the use of unsupervised clustering as a diagnostic tool for territorial analysis. Rather than providing causal explanations, the resulting spatial profiles offer a descriptive foundation that facilitates exploratory analysis and supports differentiated decision-making in public security contexts.

For illustrative purposes, a representative clustering configuration can be highlighted. In the case of San Fernando, where the Davies–Bouldin index indicated favorable cluster separation, the K-Means algorithm differentiated a small and spatially compact cluster characterized by high crime concentration from several clusters exhibiting consistently low incidence levels. The high-incidence cluster was clearly distinguishable, while low-incidence clusters displayed more dispersed patterns across the urban space. This example illustrates how territorial segmentation can reveal localized crime dynamics and enhance interpretability without introducing causal assumptions or additional quantitative complexity.

Regarding the predictive component, the stable performance of the AdaBoost model confirms the feasibility of applying supervised machine learning techniques for short-term crime anticipation. The observed recall and AUC values indicate that the model effectively discriminates between scenarios with and without crime occurrence, even under conditions of class imbalance and spatial heterogeneity. The relatively stronger performance observed for intermediate and extended temporal windows suggests that crime risk prediction benefits from incorporating recent historical information rather than relying exclusively on immediate signals.

A key methodological decision in the proposed framework is the explicit separation between territorial segmentation and predictive modeling. Cluster labels were not incorporated as explanatory variables in the supervised model, avoiding potential issues related to spatial dependence and overfitting. This design allows each analytical component to serve a distinct purpose: spatial segmentation enhances contextual understanding and interpretability, while supervised prediction focuses on operational anticipation of crime events.

Despite these strengths, several limitations should be acknowledged. First, the results depend on the quality and consistency of administrative crime records, which may be affected by underreporting or changes in data collection practices. Second, the use of internal validation metrics for clustering constrains direct comparison with alternative approaches based on external validation. Finally, the predictive model focuses on the occurrence of crime events rather than their severity or specific typologies, which represents a potential avenue for future research.

Overall, the findings support the relevance of a complementary analytical framework that integrates unsupervised and supervised learning techniques. By balancing spatial interpretability and predictive capability, the proposed approach offers a flexible and replicable solution for crime analysis and short-term anticipation in territorially heterogeneous contexts.

7. Conclusions and Future Work

This study proposed a dual analytical framework for crime analysis that integrates territorial segmentation through unsupervised learning and short-term crime prediction using supervised machine learning. Unlike approaches that address these tasks in isolation or combine them into a single analytical pipeline, the proposed design maintains a clear conceptual and methodological separation between both components, allowing each to serve a distinct analytical purpose.

Clustering-based territorial segmentation enabled the identification of differentiated spatial patterns at the neighborhood level across the analyzed municipalities, highlighting the territorial heterogeneity of crime dynamics. The use of the Davies–Bouldin index revealed that clustering quality

varies depending on local conditions, reinforcing the relevance of conducting analysis at intra-urban scales. These findings confirm the value of unsupervised clustering as a descriptive tool for spatial crime diagnosis.

The predictive component based on the AdaBoost algorithm demonstrated stable and consistent performance in anticipating crime events over short-term temporal horizons. The prioritization of recall contributed to reducing the proportion of false negatives, which is particularly important in public security applications. Results suggest that crime risk prediction benefits from incorporating recent historical information, especially when using intermediate and extended prediction windows.

Overall, the proposed framework offers a flexible and replicable approach that balances territorial interpretability and predictive capability. Its modular design facilitates adaptation to different urban contexts and data availability scenarios, making it especially relevant for settings with limited analytical resources. Future research directions include the incorporation of additional spatial variables, the exploration of external validation approaches for territorial segmentation, and the extension of the predictive model toward estimating crime severity or specific offense typologies.

Author Contributions: Conceptualization, E.R. and J.G.M.; methodology, J.G.M.; software, J.G.M.; validation, E.R. and R.P.; formal analysis, J.G.M.; investigation, J.G.M.; data curation, J.G.M.; writing—original draft preparation, J.G.M.; writing—review and editing, E.R., J.G.M., and R.P.; visualization, J.G.M.; supervision, E.R. and R.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The crime data used in this study were obtained from official administrative records provided by public security institutions. Due to legal and privacy restrictions, the raw data are not publicly available. Aggregated data and derived features used for analysis may be made available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to thank the Universidad Autónoma de Tamaulipas for providing institutional support for this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Azis, N.; Ali, Y.; Subekti, R.; Junianto, P.; Diner, L.; Suhendra, S.; Aisyah, S.; Windarto, A.P. Mapping Study Using the Unsupervised Learning Clustering Approach. *IOP Conference Series: Materials Science and Engineering* **2021**, *1088*, 012005. <https://doi.org/10.1088/1757-899X/1088/1/012005>.
2. Vulandari, R.; Laksito, W.; Aditama, D. Application of K-Means Clustering in Mapping of Central Java Crime Area. *Indonesian Journal of Applied Statistics* **2020**, *3*, 38. <https://doi.org/10.13057/ijas.v3i1.40984>.
3. Wahidah, Z.; Utari, D.T. Implementation of K-Means Algorithm to Group Provinces by Factors Influencing Criminal Acts in Indonesia. *Enthusiastic International Journal of Applied Statistics and Data Science* **2022**, *2*, 37–46. <https://doi.org/10.20885/enthusiastic.vol2.iss1.art5>.
4. Jha, G.; Ahuja, L.; Rana, A. Criminal Behaviour Analysis and Segmentation Using K-Means Clustering. In *Proceedings of the 8th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)*; 2020; pp. 1356–1360.
5. Krishnendu, S.G.; Lakshmi, P.P.; Nitha, L. Crime Analysis and Prediction Using Optimized K-Means Algorithm. In *Proceedings of the Fourth International Conference on Computing Methodologies and Communication (ICCMC)*; 2020; pp. 915–918.
6. Singh, R.; Reddy, R.; Kapoor, V.; Churi, P. K-Means Clustering Analysis of Crimes Against Women in India. *Journal of Cybersecurity and Information Management* **2020**, 5–25. <https://doi.org/10.54216/JCIM.040101>.
7. Chanchí Golondrino, G.E.; Ospina Alarcón, M.A.; Muñoz Sanabria, L.F. Cybercrime Characterization in the Department of Cundinamarca through Exploratory Analysis and Machine Learning. *Ingeniería y Competitividad* **2022**, *25*. <https://doi.org/10.25100/iyv.v25i1.11760>.

8. Mandalapu, V.; Elluri, L.; Vyas, P.; Roy, N. Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions. *IEEE Access* **2023**, *11*, 60153–60170. <https://doi.org/10.1109/ACCESS.2023.3286344>.
9. Jenga, K.; Catal, C.; Kar, G. Machine Learning in Crime Prediction. *Journal of Ambient Intelligence and Humanized Computing* **2023**, *14*, 2887–2913. <https://doi.org/10.1007/s12652-023-04530-y>.
10. Parthasarathy, S.; Lakshminarayanan, A.R. Naïve Bayes–AdaBoost Ensemble Model for Classifying Sexual Crimes. *Proceedings of the International Conference on Computing and Data Science*; 2022.
11. Arslan, H.; Doluca Horoz, A. Crime Analysis and Forecasting Using Machine Learning. *Journal of Optimization and Decision Making* **2023**, *2*, 270–275.
12. Tamir, A.; Watson, E.; Willett, B.; Hasan, Q.; Yuan, J.-S. Crime Prediction and Forecasting Using Machine Learning Algorithms. *International Journal of Computer Science and Information Technology* **2021**, *12*, 26–33.
13. Khatun, R.; Ayon, S.I.; Hossain, R.; Alam, J. Data Mining Technique to Analyse and Predict Crime Using Crime Categories and Arrest Records. *Indonesian Journal of Electrical Engineering and Computer Science* **2021**, *22*, 1052–1060. <https://doi.org/10.11591/ijeecs.v22.i2.pp1052-1060>.
14. Davies, D.L.; Bouldin, D.W. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1979**, *1*, 224–227.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.