Strategies of deep learning for crime forecasting in multiple regions

Estrategias de aprendizaje profundo para el pronóstico de la criminalidad en múltiples regiones

Martín Solís[1], Luis-Alexander Calvo-Valverde[2]

Solís, M; Calvo-Valverde, L.A. Strategies of deep learning for crime forecasting in multiple regions. *Tecnología en Marcha*. Vol. 39 Nº especial sobre Inteligencia Artificial. Febrero, 2026. Pág. 299-309.

https://doi.org/10.18845/tm.v39i5.8494

[1] Business School, Instituto Tecnológico de Costa Rica, Cartago, Costa Rica.
marsolis@itcr.ac.cr
https://orcid.org/0000-0003-4750-1198

[2] Computer Engineering School, Instituto Tecnológico de Costa Rica, Cartago, Costa Rica.
lcalvo@itcr.ac.cr
https://orcid.org/0000-0003-3802-9944

Keywords

Crime forecasting; fine-tuned models; deep learning; artificial intelligence.

Abstract

This study compares the crime prediction performance across 83 regions of fine-tuned pre-trained models versus models trained from scratch, using different strategies. The fine-tuned Lag-Llama model, using a strategy of training of a unique model that can predict any of the 83 regions was the best for monthly predictions, while the fine-tuned Lag-Llama using the strategy of training by groups of time series created with the k-means method was the best for the daily predictions. Apparently, the clustering training strategy allows the Lag-Llama to make a better fine-tuned for time series with characteristics that make them less predictable, such as nonlinearity and variability. Even though the Lag-Llama showed the best results at the general level, it is not the best model to make crime predictions for every region. There are models more suitable for some regions. Therefore, it is advisable to implement more than one model in a crime forecasting system.

Palabras clave

Pronóstico del crimen; ajuste de modelos; aprendizaje profundo; inteligencia artificial.

Resumen

Este estudio compara el desempeño en el pronóstico de delitos de 83 regiones del ajuste de modelos pre-entrenados con modelos entrenados desde cero, usando diferentes estrategias. El modelo Lag-Llama ajustado, utilizando una estrategia de entrenamiento de un modelo único que puede predecir cualquiera de las 83 regiones fue el mejor para predicciones mensuales, mientras que el Lag-Llama ajustado utilizando la estrategia de entrenamiento por grupos de series de tiempo creadas con el método k-medias fue el mejor para las predicciones diarias. Aparentemente, la estrategia de entrenamiento de agrupamiento permite que el Lag-Llama haga un mejor ajuste para series de tiempo con características que las hacen menos predecibles, como la no linealidad y la variabilidad. Si bien el Lag-llama mostró los mejores resultados a nivel general, no es el mejor modelo para hacer predicciones de delitos para todas las regiones. Hay modelos más adecuados para algunas regiones. Por lo tanto, es aconsejable implementar más de un modelo en un sistema de predicción de delitos.

Introduction

Empirical evidence has shown how the strategies derived from crime prediction led to crime reduction [1]. The forecast derived from trained models has been used to anticipate the incidence of crime, carry out predictive surveillance [2], and determine how to distribute the material and monetary resources to combat the crime. The forecasts also could be used to evaluate policy strategies, comparing the prediction with the real values. The predictions can be used as the counterfactual of an intervention. Considering the benefits that crime forecasting could have on crime reduction, it is relevant to research how to improve the predictions.

The research on crime forecasting has been focused on the generation of models based on statistical and machine learning algorithms that are trained using time series of crime occurrences. Recently, the area of forecasting has begun to adapt the architecture of powerful models created in Natural language Processing like ChatGPT, Llama, and others, to train models on many time series [3,4 5].

Since these models have learned the patterns of many time series, they can be used to make predictions of other time series: using the model as it was created (it is known as zero-shot forecasting) or updating the pre-trained model to new time series (it is known as fine-tuned).

This study analyzes and compares the performance of fine-tuned pre-trained models for time series of crime reports versus models trained from scratch. Specifically, the study compares the performance of an LSTM encoder-decoder architecture that has given the best results for predicting crimes in a previous study [6] with the fine-tuned of two models trained with several time series and architectures adapted from large language models. As far as we know, this is the first study that uses pre-trained models to make predictions of crime time series.

Crime forecasting often requires the prediction of the number of crimes for multiple areas in a region or country. Therefore, it is possible to create a model for each area of interest, create models for groups of areas with similar behaviors, or create a single model that can predict an area of interest. This study analyzes which of the three strategies is the best for the training of a forecasting model from scratch and the fine-tuned of a pre-trained model. As far as we know, this is the first study that investigates which way of training could be the best option.

In summary, this study compares the performance between the fine-tuned of pre-trained models with models training from scratch for crime forecasting and analyzes the interaction between the type of models used with three different strategies for model training.

Recent Studies for Crime Prediction

The research on crime prediction can be divided into classification methods for crime prediction and regression methods for crime prediction [7] The focus of classification is to determine if in an area a crime will occur. For example, [8] created models for the binary classification of areas where a crime will occur at a specific time. The areas came from splitting a city into cells of the same size.

On the other hand, the focus of regression is the prediction of a numerical value related to the cases of crimes that will occur in a certain time window in the future. In regression, we have found studies that are intended to predict the crime rate. This is the case of [9], who predict the future yearly crime rate of EEUU, and [10] who predict the crime rate of the most common crimes in Chicago. Most studies in the regression task are related to the prediction of the count of crime cases instead of the crime rate. This is the emphasis of this manuscript.

Recent studies on the prediction of the count of crime cases have been using different methods that could be classified as statistical methods, and machine learning methods. In relation to the statistical methods, we have found that recent studies used ARIMA [e.g 11, 12], auto.ARIMA [e.g 13], SARIMA [e.g14], and Holt-Winters [e.g 15] to create forecasting models. In these studies, the statistical model was not presented as the main proposal of the manuscript, but as a baseline for machine learning models.

In relation to machine learning methods, most of the recent studies created deep learning models. For example, [16] proposed a Seq2Seq based encoder-decoder LSTM model to predict the subsequent week of total daily crime in Chicago cities. The LSTM encoder encodes the time series sequence in one fixed-length context vector. The LSTM decoder takes the context vector to produce the output sequence prediction. [12] were also based on the encoder-decoder architecture to propose a novel paradigm of Dual-robust Enhanced Spatial-temporal Learning Network (DuroNet). The model is robust to outliers and waves because it uses local temporal context information to smooth the deviation of outliers and self-attention-based to weaken the effect of irregular waves. Based on the deep spatio-temporal 3D convolutional neural networks, [17] created a model to make predictions using crime maps. [11] applied a Nonlinear Autoregressive Neural Network to make predictions of non-linear time series.

Other deep learning models with more simple architectures were suggested, for example, a simple LSTM was used by [14, 18, 19]. [15] used simple Neural Networks (ANN) and other methods, but additionally compared different techniques based on the extraction of time series seasonality before the training of the model and the incorporation of the seasonable effect in the predictions.

There are studies that combine statistical methods and machine learning methods for crime forecasting. For example, [20] created an ensemble model that combines the predictions of several trained models like Naïve, Seasonal naïve, Decomposition (STL), Exponential smoothing (ETS), Linear model, ARIMA, Neural network, Prophet, FASSTER

A combination between a statistical model and a deep learning model is proposed by [2]. The model is composed of five parallel Exponential smoothing and Bi-LSTM for five types of crime. The exponential layer is used to extract smoothing and level components. Finally, concatenated results are forwarded to Bi-LSTM for predicting the number of crimes.

Methodology

Dataset

Our analysis relied on property crime data spanning from January 1, 2015, to May 30, 2023, compiled by the Costa Rican Judicial Research Department (OIJ). This dataset encompasses detailed information regarding the geographical and temporal attributes of each reported crime incident. The crime reports were aggregated into a spatial and temporal unit. The spatial level is composed of 83 regions into which Costa Rica is divided (these regions are called cantons). We work with two temporal levels: day and month. Therefore, we generated models to make daily and monthly forecasts for the 83 regions.

Models

At the daily level, the models were trained to predict the number of crimes that will occur in the next 10 days, and at the monthly level, the models predict the next 6 months. We used as input the crime counts of 16 past months and the values of 30 past days. We worked with the three deep learning architectures that are described below.

Lag-Llama

Lag-Llama, proposed by [3] is a decoder-only transformer that receives as input a context-length window of a time series . The context window of size C contains prior values of the time series and covariates . Thus, the vector of is concatenated with the vector of . The covariates are composed of date-time features (month, day, week, etc), and descriptive statistics of each window.

The input, known as a token, is passed through a shared linear projection layer and then through M-masked decoder layers. Finally, the output of the decode stage is the input of the last layer, known as the distribution head. This last layer uses a probability distribution as the t-student, to predict the parameters φ of the forecast distribution of the next time step. This last layer lets to estimate a range of possible outcomes instead of a punctual deterministic prediction. The range of values can be averaged to obtain the punctual forecast, but more importantly, could be used to measure the uncertainty of the estimations.

The Lag-Llama architecture was trained on 27 time series datasets from different domains. The pre-trained model showed a good performance for zero-shot forecasting, and state-of-the-art performance after fine-tuned. We used the Lag-Llama model uploaded in https://github.com/time-series-foundation-models/lag-llama

PacthTST

PathTST is a model created by [21] that outperforms other baselines. A group of n time series data is divided into different channels that share the same Transformer backbone, but the forward processes are independent. In each channel, the univariate time series is divided into M patches (split the time series into M subseries). If the patch length is equal to P, the length of the time series is equal to L and there is a stride S (non-overlapping region between two consecutive patches), the number of patches is (L-P)/S +2. The channels fed a Transformer encoder with a multi-head attention block, BatchNorm layers, and a feed-forward network with residual connections. The encoder output is passed to the last layer that generates the sequence prediction. This is a flattened layer with a linear head.

According to the authors, the patching approach offers three key advantages: it preserves local semantic details within the embedding, significantly reduces computational and memory requirements for attention maps by a quadratic factor while maintaining the same look-back window and enables the model to extend its attention span to encompass longer historical contexts. For PacthTST we used the model in https://github.com/yuqinie98/PatchTST

Encoder-Decoder LSTM

This is a model proposed by [6]. The model showed a performance better than SARIMAX and other five deep learning models for crime forecasting. The architecture is composed of an Encoder LSTM (with recurrent units and a linear activation function) + Decoder LSTM + time distributed layer (output is obtained separately for each time step). The Encoder-Decoder was trained from scratch.

Strategy of training

As we mentioned in the Introduction, three strategies of training were analyzed:

a) Individual: In this strategy, one model was generated for each of the 83 regions. For PatchTST the fine-tuned process was applied for each region independently of the other regions. For Lag-Llama, this strategy was not applied because the fine-tuned process requires two or more time series.

b) Cluster: In this strategy, we generated three groups of time series, according to the similarity between the 83 daily time series and the similarity of the 83 monthly time series. The method used to group the time series was k means with the Euclidean distance. To determine the appropriate number of clusters, the Calinski–Harabasz score values are analyzed under different cluster size configurations for monthly and daily time series. The results obtained support that one of the most suitable solutions is the selection of three clusters. The number of monthly time series by cluster were: 17, 15, 51, and for the daily time series were: 21, 15, 47. The training of the encoder-decoder and the fine-tuned of the other two architectures were generated by group. Therefore, 6 models were created for each method, 3 models for the daily time series and 3 models for the monthly time series. The assumption behind this strategy is that the learning of the models could improve if the time series has similar patterns.

c) All: In this strategy, the training and fine-tuned were generated using the 83-time series of the daily and monthly time series. Therefore, only two models were generated for each method. The assumption here is that the model will learn better from the diversity of patterns.

Experimental procedure

Cross-validation on a rolling basis with three folds was implemented to evaluate the performance of the models and the strategies. For each fold, a modified version of the metric Relative Mean Absolute Error (RelMAE) was computed. The three RelMAE were averaged for each time series. The original (RelMAE) is the division of the Mean Absolute Error of the forecasting model by that of a benchmarking method. Our benchmarking method is the estimation of the future crime counts using the average of the crime counts of the last six months and the last ten days. This is a naïve method that can be used by people who don't know about forecasting methods. Thus, the metric compares the performance of our models with a simple prediction that can be used by people who don't know about forecasting methods. We changed the original RelMAE, including the mean absolute error of the naïve prediction in the numerator instead of in the denominator, to avoid the non-definition when no crime occurred in the last six months or the last ten days.

The formula is the following:

$$\text{RelMSE\_modified} = \frac{MAE\_naive}{MAE\_forecast\_method}$$

When the value of the metric surpasses the threshold of 1, the forecasting model is useful, but when it is less than 1, it is better to use a naïve prediction to estimate future crime occurrences.

Results

Table 1 shows the performance of the model based on the median of RelMAE in the test sample. We computed the median instead of the mean because there are outliers that affect the mean. There are only 83 time series, and a single outlier can have a significant influence on the mean. Table 1 shows that the Lag-Llama model independently of the strategy of training got the best performance for the daily and monthly time series. This result proves that fine-tuned can be useful for crime forecasting. Regarding the strategy of training, the medians indicate that the best strategy is the training of the models using the crime reports for all regions. This finding suggests that deep learning architectures could learn better from the variability between regions than focusing on learning the specific behavior of each region. On the hand, Table 2 shows the performance, for the combination of type of model and strategy. According to the results, at the monthly level, the Lag-Llama with the strategy of all-training performs better than other models. At the daily level the best options are the Lag-Llama with the strategy of all-training and Lag-Llama with the strategy of training by cluster.

Table 1. Median of the RelMAE in test, by model and strategy of training.

Type	Daily	Monthly
Model		
Lag-Llama	1.07	1.10
encDec	1.01	1.04
patch	1.01	1.04
Strategy		
All (A)	1.03	1.11
Cluster (C)	1.02	1.05
Individual (I)	1.01	1.00

Table 2. Median of the RelMAE in test, combining model and strategy of training.

Model	Daily	Monthly
encDec_A	0.99	1.42
encDec_C	1.02	1.35
encDec_I	1.01	1.21
patch_A	1.02	1.36
patch_C	1.01	1.33
patch_I	1.02	1.31
Lag-Llama_A	1.07	1.44
Lag-Llama_C	1.07	1.38

Note: C: cluster strategy, A: All strategy, I=Individual

Table 2 does not show the number of time series (regions) that are best predicted for each model. Meanwhile, Table 3 presents that information in the column named Best, and the number of time series (regions) in which the model is the best by a difference larger than 0.03 in the metric RelMAE. The values in Table 3 show that Lag-Llama_A model offers the best performance across a larger number of time series. This result confirms the findings of Table 2 about which model is the best at the monthly level. At the daily level, the results complement the findings in Table 2 and allow us to conclude that Lag-Llama_C offers the best performance.

Table 3 shows, for each model, how many time series achieved their best performance with that model and how many achieved a best performance with a RelMAE at least 0.03 units higher than that of the second-best model. This table allows for analyzing whether there is a model that predominantly outperforms the others across the majority of the time series. The findings in Table 3 suggest that there is not a unique model that can be the best predictor for each region; therefore, it is relevant to consider the implementation of more than one model to improve the performance of the predictions in multiple regions. For this reason, Table 4 compares the RelMae using the best model for each region versus the implementation of the three best models. It also shows how many regions would have an acceptable RelMae value (greater than 1) with both alternatives. In this way, Table 4 illustrates the suitability of implementing all the models or selecting a subset of them for a crime prediction system. Based on the results, we consider that the best option is the implementation of the best three models because the performance is similar to the best possible alternative (implement all models), and it reduces complexity in the implementation.

Table 3. Number of time series that are best predicted for each model.

Model	Daily		Monthly	
	Best	Best>0.03	Best	Best>0.03
Lag-Llama_C	37	21	13	13
Lag-Llama_A	14	6	24	22
encDec_I	7	4	5	3
encDec_C	6	1	5	4
patch_I	6	2	3	2
patch_A	5	0	9	4
encDec_A	4	0	17	12
patch_C	4	0	7	1

Note: C: cluster strategy, A: All strategy, I=individual

Table 4. Performance of a forecasting system based on different alternatives of models.

Alternative	Daily		Monthly	
	RelMAE>1.0	median RelMAE	RelMAE>1.0	median RelMAE
Best_three _models	70	1.11	70	1.28
Best_model_for_all_regions	75	1.14	76	1.29

We estimated the median of five features of the daily and monthly time series according to the model that was the best predictor. The objective was to find factors associated with the fact that a model gave the best performance for a group of time series. We computed the median instead of the mean because there are outliers that affect the mean. The features estimated were:

- Trend. Measure the strength of the trend based on [22]. The score is between 0 and 1, being 1 the strongest strength.
- Non-linearity. Measure the non-linearity of the time series, based on [23]. Higher values represent more non-linearity.
- Coefficient of variation. Measure the variability of the time series. Higher values represent more heterogeneity.
- Percentage of zeros in the time series.
- Mean of crime reports. Measure the number of cases by period of time.

The results in tables 5 and 6 show some relevant findings. At the daily level, two characteristics show different values between models: the proportion of zeros and the coefficient of variation. For example, the Lag-Llama with the strategy of cluster training tends to be more adequate for time series with a higher proportion of zeros and higher variability. The PatchTST model with the individual and cluster strategy got the best results with time series that have less proportion of zeros and variability. At the monthly level, the features trend, non-linearity, and mean are factors descriptively associated with the performance of the time series. Lag-Llama with the cluster strategy and Encoder-Decoder with the strategy of all-training perform better with more non-linear time series that have a higher number of cases per month. The Encoder-Decoder with the individual strategy, performs better in time series with less non-linear tendency and stronger trends. Similarly, the PatchTST models can be associated with less non-linear time series, but with a weaker trend than the Encoder-Decoder.

Table 5. Median of five features according to the model that was the best predictor for the daily time series

Method	trend	zeros	nonlinear	CV	mean
encDec_I	0.22	0.42	0.01	120.89	1.04
encDec_C	0.24	0.24	0.05	96.07	1.82
encDec_A	0.25	0.26	0.04	90.64	1.68
patch_I	0.24	0.16	0.05	80.62	2.66
patch_C	0.26	0.16	0.07	81.72	2.53
patch_A	0.20	0.40	0.02	113.42	1.07
Lag-Llama_C	0.17	0.67	0.01	167.33	0.43
Lag-Llama_A	0.21	0.32	0.01	104.15	1.36

Table 6. Median of five features according to the model that was the best predictor for the monthly time series

Method	trend	zeros	nonlinear	CV	mean
encDec_I	0.56	0.00	0.15	31.92	26.77
encDec_C	0.42	0.00	0.34	35.88	26.33
encDec_A	0.57	0.00	0.41	29.33	49.28
patch_I	0.35	0.00	0.18	30.32	24.58
patch_C	0.31	0.00	0.19	35.55	13.51
patch_A	0.42	0.00	0.21	31.43	24.05
Lag-Llama_C	0.49	0.00	0.40	30.72	59.69
Lag-Llama_A	0.49	0.00	0.25	34.09	26.57

Conclusion

The results underscore the significance of embracing architectures derived from language models that have already been trained with numerous time series in order to investigate criminal behavior. In this study, we took advantage of the learning acquired by the Lag Llama models to create predictors of the crime at a daily and monthly level in 83 regions of Costa Rica. Regarding the training strategy, the results suggest that the two strategies that produced the best results were the strategy of training models with all the time series and the strategy of training by groups of time series based on their similarity. The first strategy mentioned was the best with Lag Llama for the monthly prediction. This finding suggests that deep learning architectures could learn better from the variability between regions than focusing on learning the specific behavior of each region. The strategy of training by groups of time series based on their similarity was the best for the daily prediction. The clustering training strategy allows the Lag Llama to make better fine-tuned for time series with characteristics that make them less predictable, such as nonlinearity and variability.

Even though the Lag-Llama showed the best results at the general level, it is not the best model to make crime predictions for any region. There are models more suitable for some regions. Therefore, it is advisable to implement more than one model, but not many models because we will end up with a more complex forecasting system to implement and update. Our analysis suggests three alternatives for daily forecasts and three alternatives for monthly forecasts. Similarly, crime forecast practitioners should not focus on choosing a single model based on statistics such as the mean or median of the test performance metric. It would be optimal to improve the capacity of a forecasting system for the analysis of how much is gained in the performance by implementing one or more additional models.

Future studies in the field of crime forecasting should investigate the performance of other pre-trained models that have emerged and will continue to emerge, for example, [5] and [24]. It is important to continue analyzing when and how the strategy of training by a cluster of time series is better than the strategy of training with an all-time series of interest. Finally, a pre-trained model with a time series of crime reports should be generated in the short term in order to prove its performance with zero-shot forecasting and fine-tuned. Studies with this approach have been developed in other areas like the forecasting of macroeconomic variables [25], however, in the area of crime forecasting, there is no research as far as we know.

References

[1] A. Meijer and M. Wessels, “Predictive Policing: Review of Benefits and Drawbacks,” International Journal of Public Administration, vol. 42, no. 12, pp. 1031–1039, Feb. 2019, doi: 10.1080/01900692.2019.1575664.

[2] U. M. Butt, S. Letchmunan, F. H. Hassan, and T. W. Koh, “Hybrid of deep learning and exponential smoothing for enhancing crime forecasting accuracy,” PLOS ONE, vol. 17, no. 9, p. e0274172, Sep. 2022, doi: 10.1371/journal.pone.0274172.

[3] K. Rasul, A. Ashok, A. R . Williams, A.Khorasani, G.Adamopoulos, R. Bhagwatkar, I.Rish, Lag-llama: Towards foundation models for time series forecasting, 2023. Available: *arXiv preprint arXiv:2310.08278*.

https://doi.org/10.48550/arXiv.2310.08278

[4] A. Garza, A, M. Mergenthaler-Canseco. TimeGPT-1, 2023. Available: *arXiv preprint arXiv:2310.03589*. https://doi.org/10.48550/arXiv.2310.03589

[5] G. Woo, C. Liu, A. Kumar, C. Xiong, S. Savarese, and D. Sahoo, “Unified training of universal time series forecasting transformers,” *arXiv*, 2024. Available: https://doi.org/10.48550/arXiv.2402.02592

[6] M. Solís and L.-A. Calvo-Valverde, “Deep Learning for Crime Forecasting of Multiple Regions, Considering Spatial–Temporal Correlations between Regions,” ITISE 2024, p. 4, Jun. 2024, doi: 10.3390/engproc2024068004.

[7] V. Mandalapu, L. Elluri, P. Vyas, and N. Roy, “Crime prediction using machine learning and deep learning: A systematic review and future directions,” IEEE Access, vol. 11, pp. 60153–60170, 2023, doi: 10.1109/ACCESS.2023.3286344.

[8] P. Stalidis, T. Semertzidis, and P. Daras, “Examining deep learning architectures for crime classification and prediction,” Forecasting, vol. 3, no. 4, pp. 741-762, 2021, doi: 10.3390/forecast3040046.

[9] V. Madhu, “Crime Rate Prediction Using Time Series Forecasting,” Master’s thesis, National College of Ireland, 2023.

[10] J. Austin and R. Rosenfeld, “Forecasting US Crime Rates and the Impact of Reductions in Imprisonment: 1960–2025,” CrimRxiv, 2023. [Online]. Available: https://doi.org/10.21428/cb6ab371.395a314b

[11] H. K. Sharma and R. M. Tailor, “Exploring the Potential of AI for Urban Crime Prediction in India: A Case Study of Indian Metropolitan Cities,” in 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), 2023, pp. (if known), doi: 10.1109/IC3I59117.2023.10398043.

[12] K. Hu, L. Li, J. Liu, and D. Sun, “DuroNet,” ACM Transactions on Internet Technology, vol. 21, no. 1, pp. 1–24, 2021, doi: 10.1145/3432249.

[13] S. Sharmin, F. I. Alam, A. Das, and R. Uddin, “An Investigation into Crime Forecast Using Auto ARIMA and Stacked LSTM,” in *2022 International Conference on Innovations in Science, Engineering and Technology (ICISET)*, 2022, pp. (if known), doi: 10.1109/ICISET54810.2022.9775862.

[14] N. Ibrahim, S. Wang, and B. Zhao, “Spatiotemporal crime hotspots analysis and crime occurrence prediction,” in *Advanced Data Mining and Applications: 15th International Conference, ADMA 2019, Dalian, China, November 21–23, 2019, Proceedings 15*. Springer International Publishing, 2019, pp. 579-588.

[15] M. A. Cruz-Nájera et al., “Short Time Series Forecasting: Recommended Methods and Techniques,” *Symmetry*, vol. 14, no. 6, pp. 1231-1243, 2022, doi: 10.3390/sym14061231.

[16] J. Alghamdi and Z. Huang, "Modeling daily crime events prediction using seq2seq architecture," in *Databases Theory and Applications: 32nd Australasian Database Conference, ADC 2021, Dunedin, New Zealand, January 29–February 5, 2021, Proceedings 32*. Springer International Publishing, 2021, pp. 192-203.

[17] Q. Dong, Y. Li, Z. Zheng, X. Wang, and G. Li, "ST3DNetCrime: Improved ST-3DNet Model for Crime Prediction at Fine Spatial Temporal Scales," *ISPRS International Journal of Geo-Information*, vol. 11, no. 10, p. 529, 2022, doi: 10.3390/ijgi11100529.

[18] W. Safat, S. Asghar, and S. A. Gillani, "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques," *IEEE Access*, vol. 9, pp. 70080-70094, 2021, doi: 10.1109/ACCESS.2021.3078117.

[19] A. Stec and D. Klabjan, "Forecasting crime with deep learning," *arXiv*, 2018. [Online]. Available: https://doi.org/10.48550/arXiv.1806.01486

[20] M. Ashby, "Forecasting crime trends to support police strategic decision making," *CrimRxiv*, 2023. [Online]. Available: https://doi.org/10.21428/cb6ab371.8c79f146

[21] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A time series is worth 64 words: Long-term forecasting with transformers," *arXiv*, 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2211.14730

[22] Y. Kang, R. J. Hyndman, and K. Smith-Miles, "Visualising forecasting algorithm performance using time series instance spaces," International Journal of Forecasting, vol. 33, no. 2, pp. 345–358, Apr. 2017, doi: 10.1016/j.ijforecast.2016.09.004.

[23] X. Wang, K. Smith, and R. Hyndman, "Characteristic-Based Clustering for Time Series Data," Data Mining and Knowledge Discovery, vol. 13, no. 3, pp. 335–364, May 2006, doi: 10.1007/s10618-005-0039-x.

[24] A. F. Ansari *et al*., "Chronos: Learning the language of time series," *arXiv*, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2403.07815

[25] M. Solís and L. A. Calvo-Valverde, "A Proposal of Transfer Learning for Monthly Macroeconomic Time Series Forecast," *Eng. Proc.*, vol. 39, no. 1, p. 58, 2023.

Declaración sobre uso de Inteligencia Artificial (IA)

Los autores aquí firmantes declaramos que no se utilizó ninguna herramienta de IA para la conceptualización, traducción o redacción de este artículo.