



**Biblioteca INFOTEC**

Ciudad de México, a 30 de mayo de 2025

## **VISTO BUENO DE TRABAJO TERMINAL**

**Maestría en Ciencia de Datos e Información  
(MCDI)**

**UNIDAD DE POSGRADOS  
PRESENTE**

Por medio de la presente se hace constar que el trabajo de titulación:

**"Análisis de la percepción del crimen y sus factores sociodemográficos en México"**

Desarrollado por el alumno: **Jesús Alfredo Hernández Orozco**, bajo la asesoría del **Dr. Elio Atenógenes Villaseñor García**, cumple con el formato de Biblioteca, así mismo, se ha verificado la correcta citación para la prevención del plagio; por lo cual, se expide la presente autorización para entrega en digital del proyecto terminal al que se ha hecho mención. Se hace constar que el alumno no adeuda materiales de la biblioteca de INFOTEC.

No omito mencionar, que se deberá anexar la presente autorización al inicio de la versión digital del trabajo referido, con el fin de amparar la misma.

Sin más por el momento, aprovecho la ocasión para enviar un cordial saludo.

Dr. Juan Antonio Vega Garfias  
Subgerente de Innovación Gubernamental

JAVG/jah

C.c.p. Mtra. Anely Mendoza Rosales. – Encargada de la Gerencia de Capital Humano. – Para su conocimiento.  
Jesús Alfredo Hernández Orozco. – Alumno de la Maestría en Ciencia de Datos e Información. – Para su conocimiento.



INFOTEC Centro de Investigación e Innovación en Tecnologías  
de la Información y Comunicación

Maestría en Ciencia de Datos e Información

## Análisis de la Percepción del Crimen y sus Factores Sociodemográficos en México

Jesús Alfredo Hernández Orozco

Director: Elio Atenógenes Villaseñor García

Artículo de divulgación requerido por la modalidad de titulación:  
*Excelencia Académica*

Aguascalientes, México

Marzo 2025

## Resumen

Existe evidencia de que la percepción de la gente sobre el crimen no siempre es consistente con las estadísticas reales de incidentes, y hay una tendencia a subestimar o sobreestimar la seguridad (Mitiku, 2019). A pesar de los considerables recursos desplegados, los esfuerzos policiales actuales no logran detener los delitos antes de que comiencen y, por lo tanto, tampoco protegen adecuadamente las vidas y las propiedades (Zhu et al., 2022). El objetivo de este proyecto fue investigar la relación entre la percepción del crimen y los factores demográficos asociados a la percepción de delitos en México. Para ello, se emplearon datos del Censo de Población y Vivienda 2020, que incluyen información como la etnia, nivel educativo, condiciones de vivienda, entre otros; mismos que se cruzaron con los resultados de la Encuesta Nacional de Seguridad Pública Urbana 2020. El propósito final es proporcionar una herramienta a entidades gubernamentales, como la Secretaría del Trabajo y Previsión Social, la SEP (Secretaría de Educación Pública) o Presidencia de la República para identificar situaciones de mejora y la formulación de políticas que contribuyan a reducir la delincuencia en las regiones más vulnerables.

**Palabras clave** — Random Forest, GridSearchCV, percepción del crimen, análisis socio-económico, clasificación supervisada

## 1. Introducción

La violencia y la delincuencia son situaciones que generan amenazas a los derechos humanos y afectan la construcción de la seguridad ciudadana. Estos temas en la actualidad, son algunas de las principales preocupaciones en la vida cotidiana de muchos gobiernos y habitantes, no sólo de México, sino de la región de América Latina; se trata de problemas multifactoriales con componentes psicosociales, culturales, económicos, político-institucionales y físico-espaciales (Flores Gutiérrez, 2021).

Existe evidencia de que la percepción de la gente sobre el crimen no siempre es consistente con las estadísticas reales de incidentes, y hay una tendencia a subestimar o sobreestimar la seguridad (Mitiku, 2019). A pesar de los considerables recursos desplegados, los esfuerzos policiales actuales no logran detener los delitos antes de que comiencen y, por lo tanto, tampoco protegen adecuadamente las vidas y las propiedades (Zhu et al., 2022).

De acuerdo a Kwon et al., 2021, un estudio en Kioto entre 2003 y 2004 que utilizó técnicas inteligencia artificial de agrupación de datos se utilizó para visualizar la extensión geográfica y la duración de los delitos de robo y fuga. Un proceso llamado categorización, que es una técnica de minería de datos utilizada para analizar los patrones de delincuencia. Agarwal et al., 2013 añaden que gracias a estas técnicas es posible extraer información significativa de datos a gran escala y que se pueden utilizar de manera efectiva para predecir tipos indefinidos de delitos.

Bajo este contexto, un estudio realizado por Safat et al., 2021 descubrió que los ingresos, los niveles de desempleo y otros factores socioeconómicos son predictores útiles de la delincuencia. Estos se utilizaron a lo largo de la investigación y que comparó distintas técnicas de inteligencia artificial entre las que se encuentran KNN, XGBoost, Random Forest y LSTM.

El objetivo de este trabajo es investigar la relación entre la percepción del crimen y los factores socioeconómicos y demográficos en México utilizando Random Forests, una técnica de inteligencia artificial ampliamente utilizada. Para ello, se emplearán datos del Censo de Población y Vivienda 2020 que incluyen información como la etnia, nivel educativo, condiciones de vivienda, entre otros e información del Directorio Estadístico Nacional de Unidades Económicas (DENUE) 2020 que Ofrece los datos de identificación, ubicación y actividad económica; mismos que se cruzarán con los resultados de la Encuesta Nacional de Seguridad Pública Urbana 2020.

El propósito final es proporcionar esta herramienta a entidades gubernamentales, como la Secretaría del Trabajo y Previsión Social, la SEP (Secretaría de Educación Pública)

o Presidencia de la República para identificar situaciones de mejora y la formulación de políticas que contribuyan a reducir la delincuencia en las regiones más vulnerables.

## 2. Trabajo relacionado

El objetivo del estudio de Agarwal et al., [2013](#) es examinar el uso del análisis de clústeres como una técnica alternativa e innovadora de detección de anomalías en el sistema de transferencias bancarias. El conjunto de datos sobre delitos utilizado para el análisis de delitos son delitos registrados por la policía en Inglaterra y Gales por delito y área de fuerza policial desde 1990 hasta 2011-12. Las técnicas de inteligencia artificial y herramientas utilizadas fueron k-means clustering, R, Tanagra, Weka, KNIME, Orange.

Cabargas Carvajal, [2019](#). El objetivo de este proyecto es la predicción De Robo De Vehículos Basado En Redes Neuronales Alimentadas Por Datos Espacio Temporales E Imágenes De Google Street View. Utilizaron el Banco de datos de robos de vehículos provisto por la AACH. Las técnicas de inteligencia artificial y herramientas utilizadas fueron Keras, Tensorflow, Redes Neronales Residuales.

Gonzalez y Leboulluec, [2019](#). El objetivo de este estudio es realizar un examen comparativo de los algoritmos de aprendizaje automático y buscar un modelo efectivo para predecir el número total de delitos violentos. El conjunto de datos titulado Comunidades y delincuencia proporcionado por el Repositorio de aprendizaje automático de la Universidad de California en Irvine. Las técnicas de inteligencia artificial y herramientas utilizadas fueron Multiple Linear Regression, Random Forest Regression, Neural Network Regression, and Bayesian Regression.

Walczak, [2021](#). Este artículo revisa y examina la investigación existente sobre el uso de redes neuronales para pronosticar delitos y resolver otros problemas de toma de decisiones policiales. Los datos para los modelos NN se obtuvieron de los datos de acceso abierto RMS Crime Incidents disponibles en la ciudad de Detroit, Michigan. Las técnicas de inteligencia

artificial y herramientas utilizadas fueron Multi-layer perceptron, Recurrent neural network, SOM y Convolutional Neural Networks

Abdulraheem et al., [2022](#). Este artículo propone un algoritmo de aprendizaje automático para indicar la frecuencia y el patrón de los delitos en función de los datos recopilados y para mostrar el alcance del delito en una región en particular. Los datos utilizados para esta investigación vienen de la base de datos de delitos pública de la ciudad de Buffalo, Estados Unidos. La técnica de inteligencia artificial utilizada fue Random Forest.

Safat et al., [2021](#). Esta investigación busca mejorar la precisión en la predicción y pronóstico del crimen utilizando modelos de aprendizaje automático, además de identificar patrones de crimen, tipos de delitos y zonas de alto riesgo para ayudar en la toma de decisiones y estrategias policiales. Las bases de datos utilizadas fueron los datos de crimen de las ciudades de Chicago y Los Ángeles. Las técnicas de inteligencia artificial y herramientas utilizadas fueron Logistic Regression, SVM (Support Vector Machine), Naïve Bayes, K-Nearest Neighbors (KNN), Decision Tree, Multilayer Perceptron (MLP), Random Forest, XGBoost.

Saravag y Kumar, [2024](#). El objetivo de este estudio es desarrollar un índice de criminalidad basado en diversas categorías que contribuyen directa o indirectamente a la violencia contra las mujeres. Evaluar la estabilidad y confiabilidad del índice mediante pruebas estadísticas y comparar modelos de predicción para identificar el mejor enfoque para capturar la variabilidad en los datos sobre violencia de género. Los datos utilizados para esta investigación provienen del Censo de India 2011, National Crime Records Bureau (NCRB) y sitios web oficiales de estados y distritos de India. Las técnicas de inteligencia artificial y herramientas utilizadas fueron Regresiones Logísticas Múltiples, Random Forests y Stochastic Gradient Descent.

## 3. Propuesta

### 3.1. Enfoque utilizado

Esta investigación tiene un enfoque mixto, ya que se basa en el análisis de datos numéricos de datos sociodemográficos y socioeconómicos junto con la percepción de la delincuencia (un aspecto subjetivo y cualitativo) proveniente de la Encuesta de Seguridad Encuesta Nacional de Seguridad Pública Urbana, fuentes de información provistas por el INEGI dentro su portal de datos. Por otro lado, esta busca ser descriptiva ya que busca caracterizar la percepción del crimen y los factores demográficos asociados, y explicativa porque intenta identificar las relaciones entre estos factores y las percepciones del delito. Por otro lado, se busca realizar un diseño correlacional que permitirá analizar si existen relaciones significativas entre los factores demográficos y la percepción de seguridad.

Este trabajo hace uso del modelo de Random Forests, debido a que existen numerosos estudios relacionados que abordan este problema utilizando estas técnicas. Por esta razón, se ha seleccionado esta metodología para su análisis.

Random Forests es un modelo de aprendizaje automático basado en árboles de decisión que puede capturar relaciones complejas y no lineales en los datos (Corrales Bonilla & Vinueza Naranjo, 2024). Esto es particularmente útil para problemas de clasificación donde la relación entre las variables predictoras (factores demográficos, socioeconómicos, etc.) y la variable objetivo (por ejemplo, percepción de inseguridad) no es lineal. Otro motivo por el cual se eligió este modelo es que “es robusto frente a la presencia de ruido en los datos” (Menze et al., 2009) además de que “este algoritmo puede seleccionar y clasificar variables mediante medidas de importancia, lo que facilita la identificación de las variables más relevantes para la predicción” (Chen & Ishwaran, 2012).

### 3.2. Conjunto de datos utilizado

La construcción de la base de datos se llevó a cabo mediante la integración de diversas fuentes de información provistas por el Instituto Nacional de Estadística y Geografía (INEGI). Estas fuentes incluyen censos, encuestas y registros administrativos disponibles a través del Sistema de Consulta de Datos Abiertos, abarcando aspectos demográficos, socioeconómicos y gubernamentales relevantes para el estudio. Sin embargo, el proceso de integración implicó un exhaustivo trabajo de limpieza y procesamiento de los datos. Se aplicaron técnicas de limpieza y transformación y la estandarización de formatos, con el fin de garantizar la coherencia de los datos para su posterior análisis y modelado predictivo de delitos. Las cuales fueron las siguientes:

- Censo de Población y Vivienda. Produce información sobre la dimensión, estructura y distribución espacial de la población, así como de sus principales características socioeconómicas y culturales, además de obtener la cuenta de viviendas y sus características, como: materiales de construcción, servicios, equipamiento e instalaciones en la misma (INEGI, 2024a).
- Encuesta Nacional de Seguridad Pública Urbana (ENSU). Tiene como objetivo obtener información que permita realizar estimaciones con representatividad a nivel nacional sobre la percepción de la población sobre la seguridad pública en su ciudad (INEGI, 2024c).
- Directorio Estadístico Nacional de Unidades Económicas (DENUE). Ofrece los datos de identificación, ubicación, actividad económica y tamaño de más de 5 millones de unidades económicas activas en el territorio nacional (INEGI, 2024b).



|                                 |                                               |
|---------------------------------|-----------------------------------------------|
| <b>Variables Dependientes</b>   | Percepción de la seguridad (seguro, inseguro) |
| <b>Variables Independientes</b> | Rango de edad.                                |
|                                 | Origen de nacimiento.                         |
|                                 | Etnia.                                        |
|                                 | Lenguas.                                      |
|                                 | Estado civil.                                 |
|                                 | Escolaridad.                                  |
|                                 | Creencias Religiosas.                         |
|                                 | Discapacidades.                               |
|                                 | Actividad económica.                          |
|                                 | Transporte particular.                        |
|                                 | Vivienda.                                     |
|                                 | Características de la vivienda.               |
|                                 | Servicios con los que cuenta la vivienda.     |
|                                 | Electrodomésticos.                            |
|                                 | Sistemas de información y telecomunicaciones. |
|                                 | Servicios de salud.                           |

Tabla 1: Variables dependientes e independientes.

Para el entrenamiento y evaluación del modelo, la base de datos final tuvo un tamaño de 61450 registros y fue dividida en dos subconjuntos: entrenamiento (80 %) y prueba (20 %), manteniendo la proporción original entre clases. La tabla siguiente muestra el número de instancias por clase en cada conjunto:

| Conjunto                       | Total de instancias | Clase 1 (Segura) | Clase 2 (Insegura) |
|--------------------------------|---------------------|------------------|--------------------|
| Entrenamiento ( <i>train</i> ) | 49160               | 26565            | 22595              |
| Prueba ( <i>test</i> )         | 12290               | 6801             | 5489               |

Tabla 2: Distribución de clases en los conjuntos de entrenamiento y prueba

### 3.3. Herramientas utilizadas

Este proyecto de investigación fue desarrollado utilizando un entorno basado en Python 3.8, debido a su popularidad y adopción en el análisis de datos y aprendizaje automático. Se emplearon Jupyter Notebooks como plataforma principal de desarrollo, lo que facilitó la escritura de código y documentar de manera fluida el proceso de análisis y experimentación.

Para la manipulación y el análisis de datos, se utilizó la librería Pandas y en cuanto a la construcción y evaluación de modelos de machine learning, se integró scikit-learn. Además, el trabajo incluyó el uso de PyTorch, una de las bibliotecas más robustas y flexibles para la implementación de random forests. Para la visualización y análisis de resultados, se recurrió a Matplotlib y Seaborn, herramientas indispensables que facilitaron la creación de gráficos.

### 3.4. Limitaciones metodológicas

- **Disponibilidad de los datos.** Debido a que los datos de Censo de Población y Vivienda se actualizan cada 10 años podrían no reflejar completamente el panorama actual del país, ya que los patrones de comportamiento criminal y las características socioeconómicas podrían haber cambiado.
- **Inconsistencia en la periodicidad de las fuentes.** Las bases de datos recopilan información en distintos intervalos de tiempo, lo que restringió el uso de múltiples fuentes. Por esta razón, se optó por utilizar únicamente el año 2020, ya que fue el único período en el que todas las bases de datos disponían de información completa.

- **Desagregación geográfica inconsistente.** Inicialmente se pretendían utilizar más bases de datos, pero los registros a nivel nacional, estatal, municipal y local no estuvieron alineados en todas las bases de datos, lo que limitó su uso.
- **Imputación de datos faltantes.** El utilizar técnicas de imputación para rellenar datos ausentes podría introducir sesgos o errores en el análisis, ya que los datos imputados no siempre representan con exactitud la realidad.
- **Variables socioeconómicas limitadas.** El uso de variables seleccionadas puede no capturar todos los factores que influyen en la delincuencia. Factores culturales, políticos o legales, que no se están midiendo, también podrían tener un impacto significativo en los resultados.
- **Sesgos en la recolección de datos.** Los datos del censo o encuestas podrían tener sesgos inherentes, como subrepresentación de ciertas poblaciones o áreas geográficas con alta criminalidad. Esto podría llevar a predicciones menos precisas para esas áreas.

### 3.5. Definición del modelo

Para la definición del modelo de Random Forests se empleó la técnica de búsqueda de hiperparámetros mediante GridSearchCV, lo que permitió realizar una búsqueda exhaustiva de la mejor combinación de parámetros. Los hiperparámetros considerados en el proceso de ajuste fueron:

- **n\_estimators:** [50, 100, 150] - Este parámetro controla el número de árboles en el bosque. Un mayor número de árboles generalmente mejora la precisión, pero también incrementa el tiempo de cómputo.
- **max\_depth:** [10, 20, 30, 40, 50, *None*] - La profundidad máxima de los árboles. Cuando se establece en *None*, los árboles se expanden hasta que todas las hojas

contienen menos de `min_samples_split` muestras.

- **`min_samples_split`**: [2, 4, 8, 16, 32] - El número mínimo de muestras requeridas para dividir un nodo interno.
- **`min_samples_leaf`**: [2, 4, 8, 16] - El número mínimo de muestras requeridas para estar en un nodo hoja.

Este proceso de búsqueda incluye la evaluación de diferentes combinaciones de estos hiperparámetros utilizando la técnica de validación cruzada para maximizar la capacidad del modelo para generalizar a datos no vistos.

Una vez obtenido el modelo óptimo, se procedió a analizar la importancia de las características a través de la propiedad `feature_importances_` del modelo. Esto permitió identificar los factores sociodemográficos que tienen mayor impacto en la percepción del crimen, proporcionando una interpretación de los resultados obtenidos.

## 4. Análisis

Para evaluar el rendimiento del modelo, se utilizó la métrica de Accuracy, que mide la proporción de predicciones correctas respecto al total de predicciones donde se analizó la matriz de confusión, recall y F1-score.

Como podemos apreciar en la imagen anterior, el resultado del modelo muestra un clasificador de Random Forest con una profundidad de 40 árboles con un mínimo de muestras requeridas de cuatro y 50 árboles.

Después de completar el entrenamiento, se llevó a cabo la evaluación del modelo, obteniendo una matriz de confusión y un informe de clasificación.

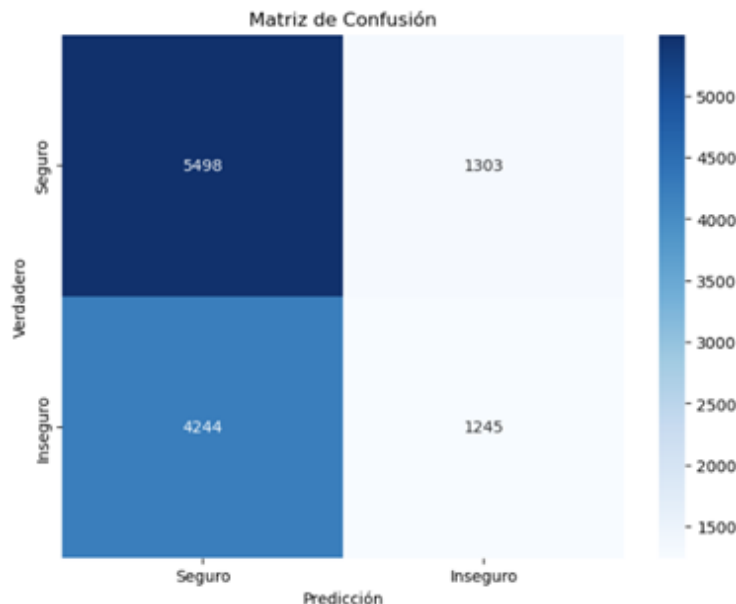


Figura 1: Matriz de Confusión del modelo hecho con Random Forests.

#### Resultados para la clase 1: Zona percibida como segura:

- Tenemos 5498 de las instancias predichas como 1 (seguro) son realmente de la clase 1 (verdaderos positivos). Indica que hay muchas predicciones correctas para la clase 1 ya que tenemos 1303 predicciones incorrectas que fueron clasificadas como 2 (inseguro) cuando deberían haber sido 1 (falsos negativos).

#### Resultados para la clase 2: Zona percibida como insegura:

- Tenemos 4244 predicciones incorrectas que fueron clasificadas como 1 cuando deberían haber sido 2 (falsos positivos) y 1245 predicciones correctas para la clase 2 (verdaderos negativos).

Si bien parece que el modelo no hace una clasificación adecuada, ya que al ver que hay muy pocas clasificaciones correctas para la Clase 1 (percepción de una zona como segura). Esto era de esperarse, debido a que no hay indicadores socioeconómicos que definan una zona como segura, ya que la pobreza, el nivel de educación o servicios por si mismos no definen una zona segura.

| Métrica   | Clase 1                                                                             | Clase 2                                                                   |
|-----------|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------|
| Precisión | 56 %                                                                                | 49 %                                                                      |
| Recall    | 81 % de los casos realmente "seguros" fueron clasificados correctamente como tales. | 66 % de los casos realmente inseguros" fueron clasificados correctamente. |
| F1-Score  | 66 %. Indica un buen rendimiento en la predicción de la zona "segura".              | 31 %. Muestra un bajo rendimiento en la predicción de la zona inseguro".  |
| Support   | 6801 casos.                                                                         | 5489 casos.                                                               |

Tabla 3: Reporte de clasificación del modelo de Random Forest

Finalmente vemos que el modelo tiene un Accuracy del 55 %. Como podemos ver, el modelo muestra un promedio simple de precisión y recall para ambas clases. De igual modo, tiene valores bajos, que indican un rendimiento desigual entre clases.

Mediante el uso de Random Forests, se analizó la importancia de las características con `feature_importances_`, lo cual proporcionó información sobre los factores sociodemográficos más relevantes en la percepción del crimen.

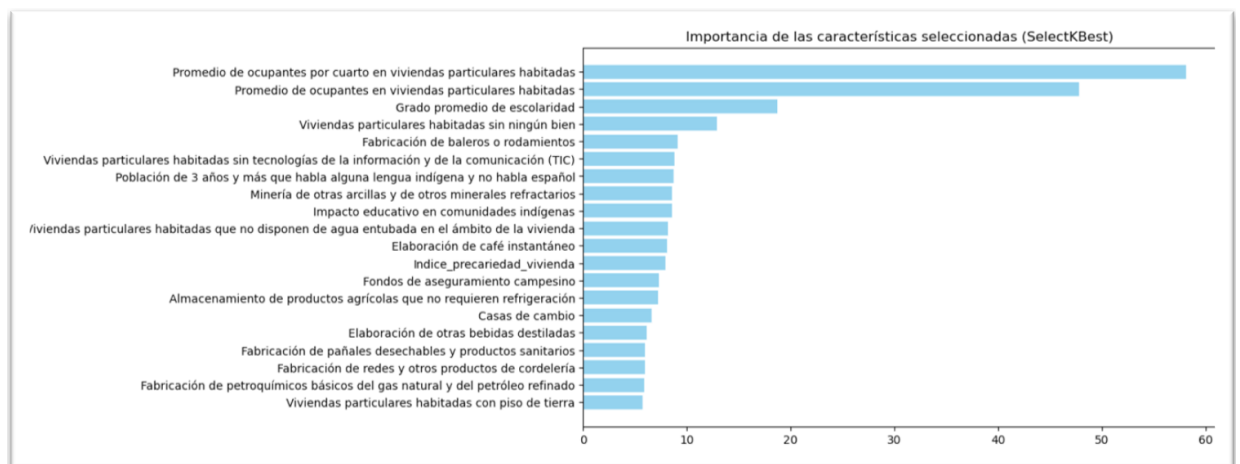


Figura 2: Gráfica de características más importantes.

Veamos algunas cosas importantes mostradas en esta gráfica:

- El número de ocupantes en las viviendas son las variables más valiosas para la identificación de la percepción de inseguridad. Esto podría reflejar condiciones de vida precarias que genera problemas sociales.

- Algo que era de esperarse es que la educación aparece como un factor relevante en la percepción de la seguridad. Una menor escolaridad podría estar asociada con mayores tasas de criminalidad o con una percepción de inseguridad más alta.
- La movilidad geográfica puede influir en la percepción del crimen, ya que las personas nacidas fuera de la entidad podrían tener experiencias o expectativas diferentes respecto a la seguridad.
- Un factor interesante fue la presencia de la religión como una característica importante para este modelo. Donde la falta de este parece indicar un vínculo con la percepción del crimen.

## 5. Conclusiones

Como se sabe, la inseguridad es un fenómeno multifactorial que, además de los factores socioeconómicos, está influido por la corrupción, la presencia del narcotráfico y otros elementos. Aunque conocer las características socioeconómicas de la población es crucial, es necesario contar con información adicional, como los niveles de delincuencia, narcotráfico y corrupción, para obtener una visión más completa.

Como se esperaba, las variables muestran que la percepción de inseguridad está fuertemente ligada a indicadores de desigualdad socioeconómica, acceso a servicios básicos y oportunidades de desarrollo. Las relaciones negativas resaltan cómo mejores condiciones educativas, tecnológicas, habitacionales y recreativas disminuyen la percepción de inseguridad, mientras que la carencia en estos aspectos la aumenta. Esto sugiere que políticas públicas enfocadas en mejorar la calidad de vida pueden ser clave para abordar la percepción de inseguridad en las comunidades.

Aunque el modelo presentó un desempeño limitado en términos de generalización, los datos obtenidos son consistentes con la realidad actual del país. Esta información podría ser útil para desarrollar políticas que contribuyan a mejorar la situación.

Por último, es importante destacar que el preprocesamiento adecuado de las variables es fundamental para la calidad de los resultados en modelos predictivos. La transformación y selección de variables, así como el manejo de valores atípicos y la normalización de datos, juegan un papel crucial en la mejora de la precisión y la interpretabilidad de los modelos.

Como trabajo futuro, se propone enriquecer el modelo mediante la integración de nuevas variables contextuales (por ejemplo, datos de movilidad) provenientes de otras fuentes oficiales que se ajusten a los niveles de desagregación geográfica y extender este análisis a otros periodos.

## Agradecimientos

*Quiero expresar mi más sincero agradecimiento a todas las personas que me han brindado su apoyo y que me dieron palabras de aliento cundo más lo necesité.*

*Estoy muy agradecido con Dios por haberme dado la posibilidad de hacer esta maestría y también por el apoyo incondicional de mi familia. En primer lugar agradezco a mi esposa, quien desde el inicio me impulsó a embarcarme en este desafiante camino. Sin duda alguna su apoyo me ayudó a cumplir mi sueño.*

*También quiero agradecer a mis padres, cuyo ejemplo ha sido mi mayor inspiración. Su confianza en mí y sus palabras de aliento fueron el combustible que necesitaba para seguir adelante incluso en los momentos más difíciles.*

*A Fabián y Carlos, mis grandes amigos, les agradezco profundamente por estar siempre presentes, por sus palabras de aliento y por recordarme constantemente que rendirse nunca fue una opción. Para mí, su amistad es un regalo invaluable que atesoro con cariño.*

*Este camino sin duda ha sido un desafío, pero también ha sido una experiencia muy gratificante y enriquecedora en mi vida.*



## Referencias

- Abdulraheem, M., Awotunde, J., Oladipo, I., Adeleke, M., Ndunagu, J., Ayantola, J., & Mohammed, A. (2022). CRIME RATE PREDICTION USING THE RANDOM FOREST ALGORITHM. *LAUTECH Journal of Engineering and Technology*, 16(2), 166-179. <https://www.laujet.com/index.php/laujet/article/view/547>
- Agarwal, J., Nagpal, R., & Sehgal, R. (2013). Crime Analysis using K-Means Clustering. *International Journal of Computer Applications*. *International Journal of Computer Applications*, 84. <https://doi.org/10.5120/14433-2579>
- Cabargas Carvajal, P. A. (2019). *Predicción De Robo De Vehículos Basado En Redes Neuronales Alimentadas Por Datos Espacio Temporales E Imágenes De Google Street View* [Tesis de maestría, Universidad de Chile].
- Chen, X., & Ishwaran, H. (2012). Random Forests for Genomic Data Analysis. *Genomics*, 99, 323-9. <https://doi.org/10.1016/j.ygeno.2012.04.003>
- Corrales Bonilla, J. I., & Vinueza Naranjo, P. G. (2024). Desarrollo de un algoritmo para la predicción de tres parámetros ambientales. *Revista Multidisciplinaria de Desarrollo Agropecuario, Tecnológico, Empresarial y Humanista*, 6. <https://doi.org/10.61236/dateh.v6i2.946>
- Flores Gutiérrez, S. (2021). Análisis espacial del delito callejero en Ciudad de México, 2018. *Quivera. Revista de Estudios Territoriales*, 23, 25-47.
- Gonzalez, L., & Leboulluec, A. (2019). Crime Prediction and Socio-Demographic Factors: A Comparative Study of Machine Learning Regression-Based Algorithms. *Journal of Applied Computer Science & Mathematics*, 13, 13.
- INEGI. (2024a). Censo de Población y Vivienda 2020. Consultado el 29 de septiembre de 2024, desde <https://www.inegi.org.mx/programas/ccpv/2020/#documentacion>
- INEGI. (2024b). Directorio Estadístico Nacional de Unidades Económicas (DENUE). Consultado el 29 de septiembre de 2024, desde <https://www.inegi.org.mx/app/mapa/denue/default.aspx>

- INEGI. (2024c). Encuesta Nacional de Seguridad Pública Urbana (ENSU). Consultado el 29 de septiembre de 2024, desde <https://www.inegi.org.mx/programas/ensu/>
- Kwon, E., Jung, S., & Lee, J. (2021). Artificial Neural Network Model Development to Predict Theft Types in Consideration of Environmental Factors. *ISPRS International Journal of Geo-Information*, 10. <https://doi.org/10.3390/ijgi10020099>
- Menze, B. H., Kelm, M., Masuch, R., Himmelreich, U., Bachert, P., Petric, W., & Hamprecht, F. A. (2009). A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *National Library of Medicine*. <https://doi.org/10.1186/1471-2105-10-213>
- Mitiku, A. (2019). *Urban Crime Characterization and Pattern Development Towards Critical Analysis & Classification* [Tesis de maestría, Bahir Dar University]. <http://hdl.handle.net/123456789/10876>
- Safat, W., Asghar, S., & Gillani, S. A. (2021). Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques. *IEEE Access*, 9, 70080-70094. <https://doi.org/10.1109/ACCESS.2021.3078117>
- Saravag, P. K., & Kumar, B. R. (2024). A Hybrid Machine Learning and Regression Approach for Validating a Multi-Dimensional Crime Index in the Context of Crime Against Women. *IEEE Access*, 12, 117955-117969. <https://doi.org/10.1109/ACCESS.2024.3439721>
- Walczak, S. (2021). Predicting Crime and Other Uses of Neural Networks in Police Decision Making. *Sec. Forensic and Legal Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.587943>
- Zhu, Q., Zhang, F., Liu, S., Wang, L., & Wang, S. (2022). Static or dynamic? Characterize and forecast the evolution of urban crime distribution. *Expert Systems with Applications*, 190. <https://doi.org/10.1016/j.eswa.2021.116115>