



Facultad de Ingeniería

ISSN: 0121-1129

ISSN: 2357-5328

revista.ingenieria@uptc.edu.co

Universidad Pedagógica y Tecnológica de Colombia  
Colombia

Ordoñez-Eraso, Hugo-Armando; Pardo-Calvache, César-Jesús; Cobos-Lozada, Carlos-Alberto  
Detección de tendencias de homicidios en Colombia usando Machine Learning  
Facultad de Ingeniería, vol. 29, núm. 54, 2020, -Marzo  
Universidad Pedagógica y Tecnológica de Colombia  
Colombia

DOI: <https://doi.org/10.19053/01211129.v29.n54.2020.11740>

Disponible en: <https://www.redalyc.org/articulo.oa?id=413962511027>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

UPEM  
redalyc.org

Sistema de Información Científica Redalyc  
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal  
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso  
abierto

# Detection of Homicide Trends in Colombia Using Machine Learning

Hugo-Armando Ordoñez-Eraso; César-Jesús Pardo-Calvache; Carlos-Alberto Cobos-Lozada

**Citación:** H.-A. Ordoñez-Eraso, C.-J. Pardo-Calvache, C.-A. Cobos-Lozada, "Detection of Homicide Trends in Colombia Using Machine Learning," *Revista Facultad de Ingeniería*, vol. 29 (54), e11740, 2020.

<https://doi.org/10.19053/01211129.v29.n54.2020.11740>

**Recibido:** Julio 12, 2020; **Aceptado:** Septiembre 13, 2020;

**Publicado:** Septiembre 14, 2020

**Derechos de reproducción:** Este es un artículo en acceso abierto distribuido bajo la licencia [CC BY](#)



**Conflicto de intereses:** Los autores declaran no tener conflicto de intereses.

# Detection of Homicide Trends in Colombia Using Machine Learning

Hugo-Armando Ordoñez-Eraso<sup>1</sup>

César-Jesús Pardo-Calvache<sup>2</sup>

Carlos-Alberto Cobos-Lozada<sup>3</sup>

## Abstract

The number of violent homicides in Latin America has grown considerably in recent decades, due to the expansion and rise of organized criminal groups in rural and urban areas of the main cities of countries such as Mexico, Colombia and Venezuela. Given their high homicide rate as a consequence of the high crime rate, these countries have been classified among the most violent in the world. According to data reported by the Crime Observatory, the National Police and the Attorney General's Office of Colombia, in 2019 there were 1,032 murders in Bogotá. This data shows a homicide rate of 14.3 per 100,000 inhabitants. From this, it is estimated that between 1960 and 2019, around 226,215 homicides were generated, which is, on average, 3,834 deaths per year. In this work a random forest-based machine learning model is presented, which allows predicting violent homicide (VH) trends in Colombia for the next 5 years. The objective of the model is to serve as an instrument to facilitate decision-making in organizations such as the Prosecutor's Office and the National Police. The model was evaluated with a dataset obtained from the Criminal, Contraventional and Operational Statistical Information System (SIEDCO in Spanish) of the Prosecutor's Office, which has 2,662,402 records of crimes committed in Colombia from 1960 to 2019.

**Keywords:** data mining; homicide; machine learning; random forest.

---

<sup>1</sup> Ph. D. Universidad del Cauca (Popayán-Cauca, Colombia). [hugoordonez@unicauca.edu.co](mailto:hugoordonez@unicauca.edu.co). ORCID: [0000-0002-3465-5617](https://orcid.org/0000-0002-3465-5617)

<sup>2</sup> Ph. D. Universidad del Cauca (Popayán-Cauca, Colombia). [cpardo@unicauca.edu.co](mailto:cpardo@unicauca.edu.co). ORCID: [0000-0002-6907-2905](https://orcid.org/0000-0002-6907-2905)

<sup>3</sup> Ph. D. Universidad del Cauca (Popayán-Cauca, Colombia). [ccobos@unicauca.edu.co](mailto:ccobos@unicauca.edu.co). ORCID: [0000-0002-6263-1911](https://orcid.org/0000-0002-6263-1911)

## **Detección de tendencias de homicidios en Colombia usando *Machine Learning***

### **Resumen**

En las últimas décadas, el número de homicidios violentos en América Latina ha crecido considerablemente debido a la ampliación y auge de grupos criminales organizados en zonas rurales y urbanas de las principales ciudades de países como México, Colombia y Venezuela. Con base en el alto índice de homicidio de estos países, consecuencia de la alta criminalidad, éstos han sido clasificados dentro de los más violentos a nivel mundial. Según datos reportados por el Observatorio del Delito de la Policía Nacional y la Fiscalía General de la Nación de Colombia, en 2019 se presentaron 1.032 asesinatos en Bogotá. Estos datos arrojan una tasa de 14,3 homicidios por cada 100.000 habitantes. A partir de esto, se estima que entre 1960 y 2019 se han generado alrededor de 226.215 homicidios, unas 3,834 muertes por año, en promedio. En este trabajo se presenta un modelo de *machine learning* basado en *random forest*, el cual permite predecir las tendencias de homicidio violento (HV) en Colombia para los próximos 5 años. El proyecto tiene como objetivo servir de instrumento para facilitar la toma de decisiones en organismos como la Fiscalía General de la Nación y la Policía Nacional. El modelo fue evaluado con un conjunto de datos generado a partir del Sistema de Información Estadístico Delincuencial, Contravencional y Operativo (SIEDCO) de la Fiscalía, el cual cuenta con 2.662.402 registros de delitos realizados en Colombia desde el año 1960 hasta 2019.

**Palabras clave:** homicidio; machine learning; minería de datos; random forest.

## **Detecção de tendências de homicídios na Colômbia usando Machine Learning**

### **Resumo**

Nas últimas décadas, o número de homicídios violentos na América Latina tem crescido consideravelmente devido à ampliação e auge de grupos criminais organizados em zonas rurais e urbanas das principais cidades de países como o México, a Colômbia e a Venezuela. Com base no alto índice de homicídio destes

países, consequência da alta criminalidade, estes têm sido classificados dentro dos mais violentos a nível mundial. Segundo dados reportados pelo Observatório do Delito da Polícia Nacional e a Fiscalía Geral da Nação da Colômbia, em 2019 apresentaram-se 1.032 assassinatos em Bogotá. Estes dados mostram uma taxa de 14,3 homicídios por cada 100.000 habitantes. A partir disto, estima-se que entre 1960 e 2019 tem sido gerado ao redor de 226.215 homicídios, umas 3,834 mortes por ano, em média. Neste trabalho apresenta-se um modelo de machine learning baseado em random forest, o qual permite predizer as tendências de homicídio violento (HV) na Colômbia para os próximos 5 anos. O projeto tem como objetivo servir de instrumento para facilitar a tomada de decisões em organismos como a Fiscalía Geral da Nação e a Polícia Nacional. O modelo foi avaliado com um conjunto de dados gerado a partir do Sistema de Informação Estatístico Delinquencial, Contravencional e Operativo (SIEDCO) da Fiscalía, o qual conta com 2.662.402 registros de delitos realizados na Colômbia desde o ano 1960 até 2019. **Palavras chave:** homicídio; machine learning; mineração de dados; random forest.

## I. INTRODUCCIÓN

A nivel mundial, la alta y creciente tasa de homicidios en las últimas décadas llama la atención de los gobiernos. América Latina no es la excepción, según el informe de la ONU en 2019 [1-2], la cifra de homicidios es de 6.7 personas asesinadas por cada 100.000 habitantes [3]. El papel en este tema de la política judicial, la fiscalía y el gobierno de cada país se centra en enfocar toda su atención para el control del crimen violento [4].

En Colombia, un país que se encuentra en conflicto interno desde 1964, el tema del homicidio violento (HV) es trascendental. Se estima que, a partir de 1960 y hasta 2019, se han generado alrededor de 226.215 muertes violentas, lo que significa, en promedio, 3.834 muertes violentas por año, hecho que alerta sobre la necesidad de implementar medidas urgentes para controlar el incremento de la violencia letal en varios territorios del país. Para ello, el gobierno ha incrementado la operatividad de la Fuerza Pública enviando fuerzas militares especiales a territorios que lo requieren. También ha reforzado otros con el objetivo de controlar el aumento del homicidio y el accionar de grupos ilegales como las disidencias de las Fuerzas Armadas Revolucionarias de Colombia (FARC) y de las Autodefensas Unidas de Colombia (AUC), frentes del Ejército de Liberación Nacional (ELN), delincuencia organizada en Bandas Criminales (BACRIM), entre otros.

Por otra parte, en su lucha por el control social y territorial, los agentes del conflicto y actores armados han intimidado a la población civil al punto de que prácticas violentas han trasladado el conflicto armando de las zonas rurales a las zonas urbanas de las ciudades [5]. En ese sentido, un nuevo ordenamiento socio-espacial, amparado por el control social de grupos y estructuras criminales armadas al margen de la ley, ha aparecido en las zonas urbanas y rurales del territorio nacional.

En Colombia, en el año 2007, la tasa de homicidios se situaba en 25.4 muertes por cada 100.000 habitantes. En el 2014 [6], con el fin de disminuir la violencia y la tasa de homicidios, el Gobierno Nacional, la Policía Nacional y la Fiscalía General de la Nación implementaron programas de seguridad a través de inversión social y tecnología, esta última con la incorporación de cámaras de video vigilancia en las ciudades, que han permitido reforzar la seguridad en las zonas de mayor afluencia delictiva.

Además, para la reducción de delitos, también se han implementado los grupos especiales contra homicidios, los cuales tienen por objetivo responder rápidamente a diversas situaciones que pueden llevar a que se cometa un homicidio. Dichos grupos especiales se crearon con el objetivo de identificar jefes del crimen organizado, desarticular sus estructuras, y reducir las actividades ilícitas que realizan, así como la impunidad. Esto se logra a través de la realización de investigaciones con mayor detalle y profundidad, y de manera articulada entre Fiscalía, Policía y Alcaldías.

Para complementar el trabajo realizado en estos programas, la Fiscalía y la Policía cuentan con un Sistema de Información Estadístico Delincuencial, Contravencional y Operativo (SIEDCO) en el cual se establecieron criterios de clasificación de los delitos para registrar aquellos reportados en el territorio nacional. Con los datos de SIEDCO, se ha logrado obtener una adecuada descripción de las circunstancias en las que se desarrollan las conductas criminales, y se han realizado investigaciones que han permitido conocer indicadores y estadísticas de los delitos. Tal es el caso de [7], que presenta estadísticas del homicidio en las zonas rurales y urbanas de Colombia, un aporte importante que es mejorable utilizando técnicas de analítica de datos a partir de *machine learning* [8].

Con base en lo anterior, en este trabajo se propone un modelo de *machine learning* basado en el clasificador de Bosques Aleatorios (conocido en inglés como *random forest* [9-10]), que permite encontrar predictores de comportamientos futuros de hechos delictivos de homicidio en Colombia. El objetivo de este modelo es servir de base para los entes gubernamentales en la toma de decisiones, a partir de los datos registrados en sus sistemas de información. Para lograr esto, se analizó un conjunto de datos (*dataset*) obtenido de SIEDCO, el cual cuenta con 2.662.402 registros de delitos realizados en Colombia desde el año 1960 hasta 2019. Los resultados obtenidos por el modelo propuesto se compararon con los obtenidos por otros tres modelos de regresión para predicción: Regresión lineal, Maquinas de vectores soporte para la regresión (SRV) y KNN regresión. Los resultados obtenidos son alentadores y muestran que los homicidios en Colombia tienden a disminuir.

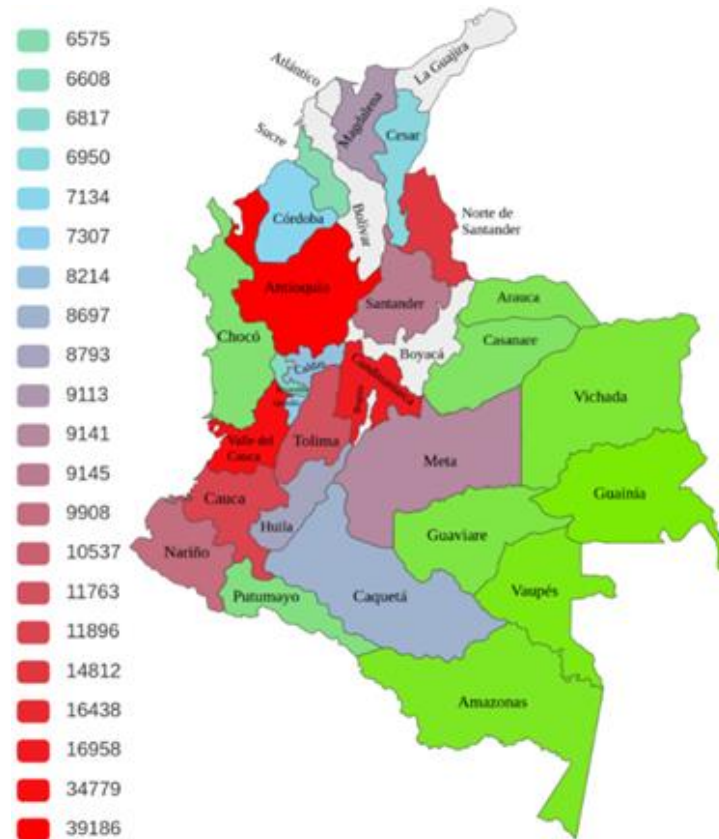
El resto del artículo se encuentra organizado de la siguiente forma, la Sección 2 presenta los trabajos relacionados, la Sección 3 presenta el modelo propuesto y la descripción de cada uno de sus componentes, la Sección 4 describe los resultados obtenidos después de implementar y probar el modelo y, finalmente, la Sección 5 presenta las conclusiones y el trabajo futuro.

## II. MÉTODOS

Según Regoeczi y T. D. Miethe [11], el homicidio puede definirse como “supresión por conducta del agente (agresor), de una vida humana (tipicidad), sin justificación jurídicamente atendible (antijuridicidad), en forma intencional o dolosa, o con culpa o preterintencional (culpabilidad)”. El homicidio es uno de los delitos más graves ya que atenta contra el bien máximo protegido que es la vida. En Colombia, la ola de homicidios que aflige al país desde hace varias décadas se ha conocido como La Violencia. Si bien, la llamada Violencia en Colombia comenzó mucho antes del año 1948 [12], la muerte de Jorge Eliécer Gaitán fue un detonante para que se generara el conflicto civil que, por más de 10 años, sumergió al país en un baño de sangre sin precedentes.



Liberales y conservadores (los dos partidos políticos dominantes de la época) empezaron una lucha frontal que dejó cerca de 300.000 muertos, lo que ocasionó el desplazamiento interno de millones de personas a las grandes ciudades del país [13]. Esta etapa marcó también el inicio de las guerrillas en el país y puso de manifiesto las más tremendas formas de violación de los Derechos Humanos. Los asesinatos de líderes también fueron recurrentes en varios departamentos a nivel nacional. Por otra parte, las organizaciones al margen de la ley, o crimen organizado, han potencializado la violencia en las zonas urbanas, lo que ha generado homicidios sistemáticos de personas involucradas y no involucradas con este tipo de organizaciones [14]. En este sentido, el conflicto armado y las organizaciones delictivas han generado muerte por todo el territorio nacional, tal como se demuestra en la Figura 1; se destacan algunos departamentos como Antioquia, donde es posible apreciar que de 1940 a 2019 han ocurrido 39.186 homicidios, de la misma forma, en el Valle del Cauca han ocurrido 34.779 muertes violentas. Como se puede observar en la Figura 1, en la mayoría de los departamentos del eje central de Colombia las cifras también son alarmantes [15]. La violencia en Colombia no distingue edad o condición social, la Tabla 1 muestra las cifras de personas asesinadas entre 1965 y 2019 [16]. En estos datos se aprecia que los adultos entre 29 y 59 años son las personas que con mayor frecuencia son asesinadas, debido a que muchos son actores activos del conflicto armado, perteneciendo a grupos de guerrilla o paramilitares que actúan en sectores rurales. Además, están las personas que pertenecen a bandas de crimen organizado enfocados en negocios ilícitos, como el narcotráfico, que operan en el sector urbano de las grandes ciudades [6-17]. Por otra parte, la población entre 18 y 28 años representa una de las cifras más altas en cuanto a homicidios, debido a que en este rango de edad algunos jóvenes ingresan a pandillas o bandas de crimen organizado por falta de oportunidades laborales. Una cifra que llama la atención es la de primera infancia, entre los 0 y 5 años, homicidios que en su mayoría se relacionan con venganzas entre los actores del conflicto, o con atentados terroristas que impactan en la vida de la población inmersa en problemas de seguridad.



**Fig. 1.** Número de homicidios por departamento en Colombia.

**Tabla 1.** Cifras de homicidios en Colombia según su edad

Grupo edad	Rango edad	Numero homicidios
Primera infancia	0 a 5	7787
Infancia	6 a 11	4441
Pre-adolescente	12 a 13	2436
Adolescente	14 a 17	13739
Juventud	18 a 28	49033
Adultez	29 a 59	65867
Adulto mayor	60 o más	21809

Del número de personas asesinadas entre 1965 y 2019, se puede apreciar que el género masculino tiene un 76.45% de los homicidios reportados, mientras que el género femenino representa el 23.74%. Esto se debe a que el género masculino es un actor de mayor actividad en el conflicto armado y conforma mayormente bandas criminales, guerrilla, o grupos paramilitares.

La disminución de los homicidios en los últimos años tiene relación con el aumento en el número de capturas. La Tabla 2 muestra el número de capturas por año, donde se observa que es mayor el número de los no capturados. Por ejemplo, de los 233.957 homicidios ocurridos entre 1965 y 2019, en el 85.17%, es decir 200.345 homicidios, no hay personas capturadas (homicidios en impunidad o sin responsables), mientras que solo el 14.82% (3.3612) de los homicidas fueron capturados.

**Tabla 2.** Datos de personas capturadas por actos de homicidio.

<b>Año</b>	<b>Capturada</b>	<b>Sin captura</b>	<b>Año</b>	<b>Capturada</b>	<b>Sin captura</b>
1965	1	0	1999	2.538	145
1978	1	0	2000	4.053	171
1979	3	0	2001	4.946	209
1980	14	0	2002	5.358	270
1981	21	0	2003	5.129	229
1982	36	0	2004	4.890	248
1983	27	0	2005	6.393	549
1984	32	2	2006	8.893	1.004
1985	108	0	2007	15.898	1.672
1986	160	1	2008	17.838	2.117
1987	215	8	2009	17.924	2.593
1988	293	7	2010	17.753	2.616
1989	231	21	2011	17.945	2.851
1990	273	8	2012	17.773	3.569
1991	342	12	2013	16.736	3.994
1992	376	7	2014	16.043	4.290
1993	469	4	2015	15.140	4.128
1994	486	20	2016	14.913	4.214
1995	528	13	2017	14.917	4.130
1996	898	80	2018	8.300	2.115

Año	Capturada	Sin captura	Año	Capturada	Sin captura
1997	1.322	68	2019	6.993	1.664
1998	1.458	74	Totales	24.7667	43.103

La Tabla 3 muestra los datos de las personas condenadas por los actos de homicidio de 1965 a 2019. Similar a los datos de la Tabla 2, el 87.25% (204.345 homicidios) no tiene condenados, esta cifra es un poco mayor a la cifra de los capturados; esto se debe a que en algunos homicidios se captura al homicida, pero no se puede comprobar su participación en los hechos y queda en libertad, permitiendo en muchos casos que cometa otros homicidios.

Los datos muestran, además, que la tasa de homicidios en las zonas rurales y urbanas tiene un enorme impacto que es catalogado como una crisis de salud pública. Como apoyo para la toma de decisiones, algunos departamentos de policía del mundo, por ejemplo el de Nueva York [18], han estado utilizando el análisis estadístico para identificar lugares de alta incidencia criminalística y responder a patrones de comportamiento criminal. En este sentido, el presente trabajo toma importancia, ya que se presenta un modelo de *machine learning* para analizar los datos y predecir la tendencia de aumento o disminución de la tasa de homicidios en Colombia y sus principales ciudades.

**Tabla 3.** Número de personas condenadas por actos de homicidio.

Año	Sin condena	Condenado	Año	Sin condena	Condenado
1965	1	0	1999	2411	272
1978	1	0	2000	3778	446
1979	3	0	2001	4547	608
1980	14	0	2002	4858	770
1981	21	0	2003	4804	554
1982	36	0	2004	4703	435
1983	27	0	2005	5892	1050
1984	34	0	2006	8237	1660
1985	108	0	2007	15030	2540
1986	156	5	2008	17058	2897
1987	213	10	2009	17291	3226

Año	Sin condena	Condenado	Año	Sin condena	Condenado
1988	247	53	2010	17407	2962
1989	227	25	2011	17736	3060
1990	253	28	2012	18358	2984
1991	335	19	2013	18164	2566
1992	381	2	2014	17694	2639
1993	438	35	2015	16793	2475
1994	459	47	2016	16859	2268
1995	511	30	2017	17148	1899
1996	860	118	2018	9695	720
1997	1222	168	2019	8398	259
1998	1317	215			

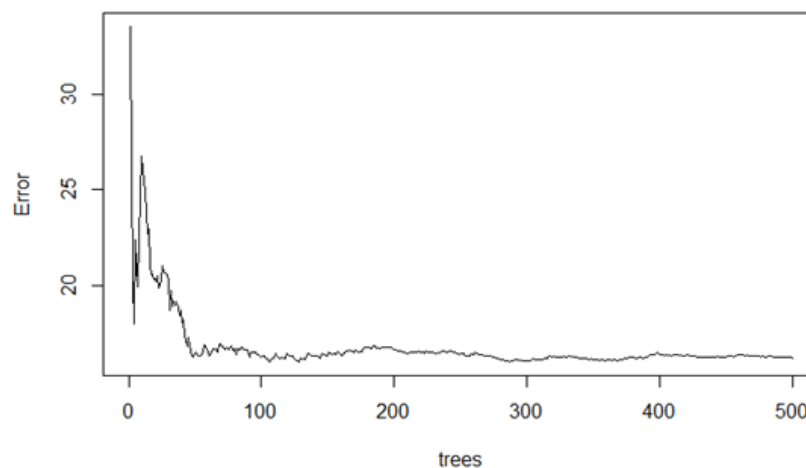
### **A. Modelo propuesto**

El modelo propuesto se basa en un algoritmo de aprendizaje supervisado, en este caso *random forest*, para un problema de regresión; el modelo utiliza para su entrenamiento un set de datos tomado del sistema de información de la Fiscalía General de la Nacional de Colombia, el cual cuenta con un total de 233.957 registros del delito de homicidio realizados en Colombia desde 1965 hasta 2019. A continuación, se describen los elementos del modelo.

El *dataset* fue filtrado y contuvo el mismo número de columnas que el original (14), entre las cuales se encuentran los siguientes atributos: conexo, estado\_noticia, etapa, anio\_hecho, ley, país, departamento, municipio, seccional, delito, captura, imputación, condena y atipicidad\_inexistencia. El proceso de minado se basó en las etapas de CRIPS-DM e inició con un análisis exploratorio de datos con el fin de obtener un entendimiento de los datos. En este proceso se removieron columnas o variables que no aportaban a la solución, se eliminaron datos duplicados, los valores faltantes fueron completados con el promedio entre el valor anterior y el siguiente en cada columna, y después se eliminaron aquellos registros que aún contenían valores nulos. Por último, en el caso de la regresión, los datos originales se normalizaron con el método Min-Max transformando los valores en un rango entre cero y uno. Se analizaron las distribuciones de las variables, los patrones que presentaban, y se identificó cómo se relacionaban las variables entre sí.

Seguidamente, se eliminaron algunas variables tales como: conexo, estado\_noticia, ley, pais, seccional, imputacion, atipicidad\_inexistencia. Además, se crearon nuevas variables como: numero-homicidios-por-ano, numero-homicidios-genero, numero-homicidios-departamento, numero-homicidios-edad, numero-capturados, numero-condenados.

Una vez filtrado el *dataset*, se definieron los hiper parámetros del modelo para que fuera afinado, tal como se muestra en la Figura 2. Los valores de hiper parámetros que mejores resultados arrojaron fueron los de número de árboles. Como se puede notar, se ejecutó el modelo hasta con 500 árboles, donde 110 es el número de árboles que mejores resultados obtiene, ya que minimiza el error a 0,67. Por otra parte, el número de variables aleatorias tomadas es 6, las cuales fueron creadas en el análisis exploratorio de datos. El número de muestras sobre las cuales se entrenó el modelo fue el 70% de los registros del *dataset*. El mínimo número de muestras dentro de los nodos terminales fue 10 y el máximo 5. Estos parámetros permiten evidenciar que, a partir de los 200 árboles, el modelo inicia a estabilizar el valor del error, y el tiempo de ejecución aumenta.



**Fig. 2.** Afinamiento de los hiper parámetros.

### III. RESULTADOS

Una vez filtrado el *dataset*, se identificaron a través de análisis de correlación de Pearson las variables que tienen mejor correlación para trabajar. Como se puede observar en la Figura 3, las variables numero\_capturados con numero\_condenados

tienen una correlación del 100% que permite identificar una relación estadística fuerte y directa. Por otra parte, se puede observar que la correlación de las variables `numero_capturados` con `numero-homicidios-por-anio` es de 0.68, es decir 68%, mientras que la correlación de `numero_condenados` con `numero_homicidios_genero` es del 96%. Estos números significan que la correlación es positiva, es decir, que las variables se correlacionan directamente.

Una vez analizadas las relaciones entre las variables, como primera instancia, se evaluó el modelo propuesto frente a tres modelos del estado del arte, estos fueron: Regresión lineal con máquinas de vectores soporte (SVR), Regresión lineal normal y Regresión lineal con K vecinos más cercanos (KNN). Para todos los modelos se realizó la predicción a 5 años, es decir, se intentó establecer cómo sería la tendencia de los homicidios hasta el año 2025 en la variable `numero_homicidios_anio` frente a `anio_hecho`. Para que la evaluación tuviera mayor coherencia, con el modelo de SVR se evaluaron las funciones *kernel*, las cuales permiten convertir lo que sería un problema de regresión no lineal en el espacio dimensional original, a un sencillo problema de regresión lineal en un espacio dimensional mayor. Las funciones evaluadas fueron radial, polinomial y lineal. La Figura 4 permite observar que el modelo de SVR implementando la función *kernel* radial (línea azul) obtiene los mejores resultados, es decir, que las predicciones en la regresión se parecen más a los datos (línea negra) o guardan más cercanía. Por lo anterior, el modelo de SVR con la función radial fue tomada para la comparación de resultados frente el modelo propuesto.

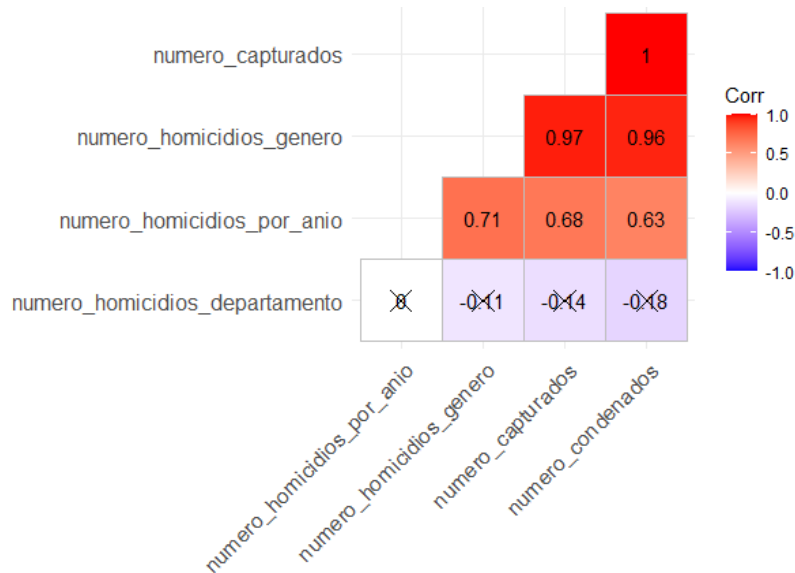


Fig. 3. Correlación de Pearson entre variables.

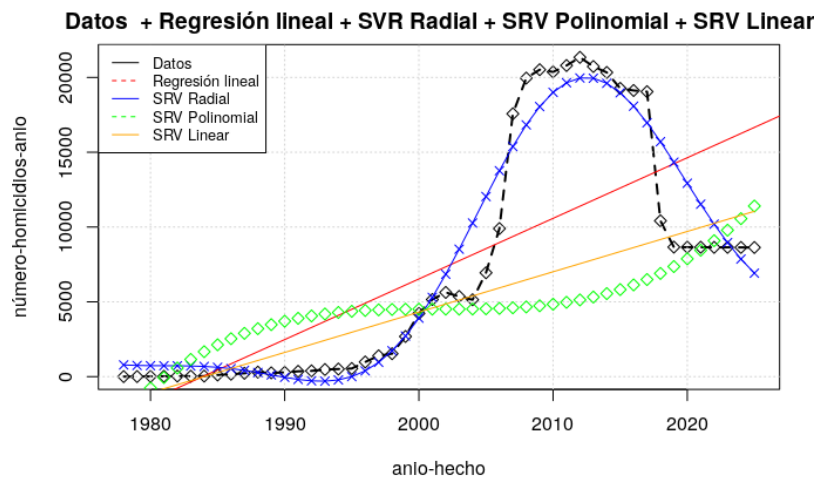
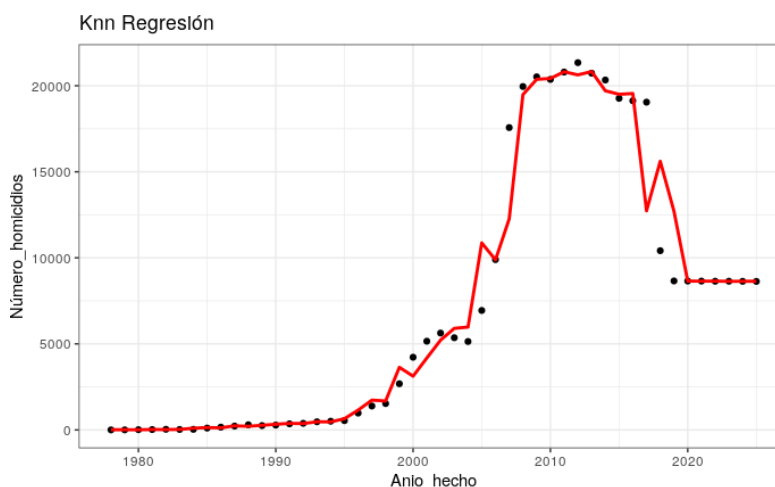


Fig. 4. Evaluación de funciones *kernel*.

Una vez seleccionado el modelo de SRV, se realizó el proceso de predicción con el modelo de KNN, la Figura 5 muestra que el modelo tiene buena predicción (línea roja), ya que guarda cercanía o similitud con los datos (puntos negros). En la predicción se puede observar que solo unos datos se dispersaron entre los años 2017 y 2020, debido a que entre los años 2019 y 2020, se predice un valor mucho más elevado que los datos originales.



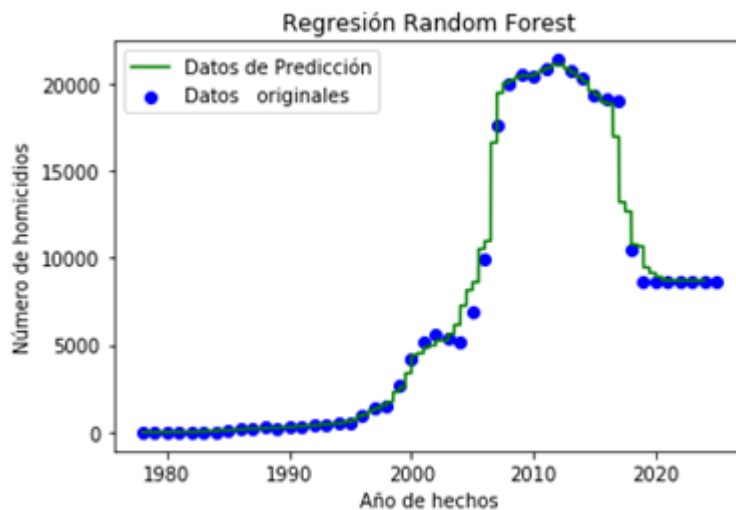


**Fig. 5.** Predicción con el modelo KNN.

Posteriormente, se evaluó el modelo propuesto, en la Figura 6 se puede observar que el modelo predice de manera armónica los valores de la variable dependiente (numero\_homicidios), ya que sus valores guardan mayor cercanía y no están muy dispersos en la regresión, esto se debe a que los árboles generados por *random forest* minimizan la distancia entre cada punto de la variable dependiente y los valores logrados en el índice Gini o entropía, en este caso, los valores de la variable hechos-año. El modelo predice que el punto más alto es en el año 2012, con un valor promedio de 21.342 homicidios, pero a medida que aumenta el valor de la variable independiente, los valores de la variable dependiente tienden a bajar, prediciendo así que para el año 2025 el número de homicidios disminuirá en Colombia; de la misma manera, para el 2021 disminuirá a 8.649 homicidios, y para el 2025 disminuirá a 843 homicidios. Esto se puede predecir ya que actores como el Estado, la Policía Nacional y la Fiscalía General de la Nación han implementado programas de convivencia, los cuales están acompañados de herramientas tecnológicas como las cámaras de video vigilancia, además, el acuerdo de paz con la guerrilla de las FARC hace que el conflicto armado disminuya sin confrontación, hecho que disminuye notablemente los homicidios en Colombia.

Una vez ejecutados los modelos y hechas las predicciones, se evaluó qué tan eficiente es el proceso de predicción, para esto se utilizaron las siguientes métricas:

- RMSE, una medida de la desviación estándar de los residuos (errores de predicción). Los residuos son una medida de qué tan lejos están los puntos de datos de la línea de regresión. RMSE, en resumen, mide qué tan concentrados están los datos alrededor de la línea de mejor ajuste.
- R2 mide coeficiente de determinación, para esta métrica los mejores valores son los que se acercan a 1 y puede ser negativa. Es decir, qué tan parecido es el valor predicho al valor esperado de  $y$ .



**Fig. 6.** Resultado de la predicción basados en el modelo propuesto.

Como se observa en la Tabla 5, los valores más bajos de RMSE los obtiene el modelo propuesto de RF, esto quiere decir que los valores predichos están mucho más cerca de los datos ideales o reales, ya que el error cuadrático medio de la solución es menor. En este sentido, RF ajustado obtuvo mejores valores que distan en 679.53 frente al modelo de regresión como KNN que logra 1391.97, es decir casi el 100% más; por otra parte, la diferencia de RMSE es bastante alta con relación a SRV y Regresión lineal.

En relación al R2, RF obtiene el valor de mayor cercanía a 1, esto quiere decir que el valor obtenido de coeficiente de determinación permite evidenciar que los valores predichos de  $y$  (numero\_de\_homicidios) son muy parecidos al valor esperado. En los demás modelos, por otra parte, KNN logra 0.824, es decir, el 15% menos de

similitud entre los valores predichos con el valor esperado; de la misma forma, SRV logra el 22% menos, alejándose un poco más que KNN de los valores esperados.

**Tabla 5.** Resultados de las métricas utilizadas.

Métrica/modelo	Random forest	knn	SRV	Regresión lineal
RMSE	<b>679.53</b>	1391.97	2095.76	5279.81
2	<b>0.976</b>	0,824	0,752	0.530

#### IV. DISCUSIÓN Y CONCLUSIONES

En este artículo se presentó el uso de un modelo de *machine learning* basado en *random forest* (RF) para predecir la tendencia del delito de homicidio desde el año 2020 hasta el 2025 en Colombia. Según los datos arrojados por el modelo, hay una tendencia a la baja en el número de homicidios para el año 2025, desde su punto más alto en el 2012, que reporta 21.342 homicidios, a solo 8.631, es decir reducción de 12.717 homicidios, una disminución del 60%.

El modelo permite obtener menores valores de RMSE y R2, lo que significa que, en el modelo propuesto de RF, los valores predichos están mucho más cerca de los datos ideales o reales. Asimismo, es posible observar que el modelo permite predecir que, con el paso del tiempo, la cantidad de homicidios irá disminuyendo, por lo que se puede pensar que las políticas estatales que se han implementado en conjunto con los entes de policía y fiscalía están funcionando. La infraestructura de video vigilancia que hace años se ha estado instalando en las principales ciudades de Colombia, acompañada con el modelo de policías por cuadrantes y la metodología de puntos críticos, ha permitido acciones preventivas y de control en cada una de las ciudades del país.

De acuerdo con el último informe de Medicina Legal, entre el año 2004 y 2019, se tuvo una reducción considerable de los homicidios en el marco del conflicto armado, lo que muestra los efectos positivos de las negociaciones de paz entre el Gobierno y los dos principales actores de grupos armados: las AUC y, ahora, la guerrilla de las FARC.

Como trabajo futuro, se espera evaluar el modelo para predecir el comportamiento de otros tipos de delitos como el desplazamiento forzado, el secuestro y el acceso carnal violento, delitos que son muy preocupantes para la sociedad y para los organismos gubernamentales como la policía y la fiscalía. Por otra parte, se espera evaluar otros clasificadores como redes neuronales, además de buscar otros *datasets* de acceso abierto para la validación del modelo propuesto.

## CONTRIBUCIÓN DE LOS AUTORES

**Hugo Armando Ordoñez Erazo:** Investigación, Análisis formal, Definición y del modelo, Implementación, Validación del modelo.

**César Jesús Pardo Calvache:** Investigación, Supervisión, Metodología, Validación, Escritura - revisión y edición.

**Carlos Alberto Cobos Lozada:** Investigación, Supervisión, Metodología, Validación, Escritura -revisión y edición.

## AGRADECIMIENTOS

Los autores agradecen el aporte de la Universidad del Cauca, donde se desempeñan actualmente como Profesores del programa de Ingeniería de Sistemas.

## REFERENCIAS

- [1] UNODOC, Global Study on Homicide - Gender-related killing of women and girls. In Unodoc, 2019. <https://doi.org/10.1023/B:0000037731.28786.e3>
- [2] A. Vazsonyi, J. Wittekind, L. Belliston, and T. Loh, *Global Study on Homicide - Homicide trends, patterns and criminal justice response*, 2019. <https://doi.org/10.1023/B:JOQC.0000037731.28786.e3>
- [3] M. M. Rogers, and J. E. Storey, "Elder homicide: A systematic literature review," *Aggression and Violent Behavior*, vol. 48, pp. 141-151, 2019. <https://doi.org/10.1016/j.avb.2019.08.008>
- [4] F. Li, S. Liu, X. Lu, Y. Ou, and P. S. F. Yip, "Application of the injury scales in homicides," *Forensic Science International*, vol. 292, pp. 83-89, 2018. <https://doi.org/10.1016/j.forsciint.2018.09.010>
- [5] M. C. Ingram, and M. Marchesini da Costa, "Political geography of violence: Municipal politics and homicide in Brazil," *World Dev.*, vol. 124, 2019. <https://doi.org/10.1016/j.worlddev.2019.06.016>
- [6] O. Fals Borda, "La Violencia en Colombia," *Entornos*, vol. 29 (2), p. 27, 2016. <https://doi.org/10.25054/01247905.1260>
- [7] M. Neira, and N. H. Martínez, *Documentos de Política Pública y Armas y homicidios*, Bogotá D. C.:

Ministerio de Salud y Protección Social, 2019.

- [8] D. E. Goin, K. E. Rudolph, and J. Ahern, "Predictors of firearm violence in urban communities: A machine-learning approach," *Health & Place*, vol. 51, pp. 61-67, 2018.  
<https://doi.org/10.1016/j.healthplace.2018.02.013>
- [9] R. Katuwal, P. N. Suganthan, and L. Zhang, "Heterogeneous oblique random forest," *Pattern Recognition*, vol. 99, e107078, 2020. <https://doi.org/10.1016/j.patcog.2019.107078>
- [10] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," *Expert Systems with Applications*, vol. 134, pp. 93-101, 2019.  
<http://doi.org/10.1016/j.eswa.2019.05.028>
- [11] W. C. Regoeczi, and T. D. Miethe, "Homicide," in *International Encyclopedia of the Social & Behavioral Sciences*, 2015.
- [12] C. Echandía, "Dimensión regional del homicidio en Colombia," *Coyuntura Social*, vol. 17, pp. 89-103, 1997.
- [13] F. Humberto, S. Murillo, J. Chica, A. Rodríguez, and G. De Cortázar, "The spatial heterogeneity of factors of femicide: The case of Antioquia- Colombia," *Applied Geography*, vol. 92, pp. 63-73, 2018.  
<https://doi.org/10.1016/j.apgeog.2018.01.006>
- [14] F. J. Escobedo, N. Clerici, C. L. Staudhammer, A. Feged-rivadeneira, J. Camilo, and G. Tovar, "Land Use Policy Trees and Crime in Bogota , Colombia : Is the link an ecosystem disservice or," *Land use policy*, vol. 78, pp. 583-592, 2018. <https://doi.org/10.1016/j.landusepol.2018.07.029>
- [15] A. M. Pardo-Monta, "Violence in Colombia and Mexico: trend and impact on life expectancy of homicide mortality between 1998 and 2015," *Public Health*, vol. 3, pp. 1-8, 2018.  
<https://doi.org/10.1016/j.puhe.2018.06.001>
- [16] N. M. Valencia, "Proceso de reconstrucción de memoria," *Revista de Estudios Sociales*, vol. 9, pp. 13-38, 2013.
- [17] L. C. Cartagena, "Los estudios de la violencia en Colombia antes de la violentología," *Diálogos Revista Electrónica*, vol. 17 (1), pp. 1-30, 2015. <https://doi.org/10.15517/dre.v17i1.18103>
- [18] K. Tardiff, P. M. Marzuk, K. Lowell, L. Portera, and A. C. Leon, "A study of drug abuse and other causes of homicide in New York," *Journal of Criminal Justice*, vol. 30 (4), pp. 317-325, 2002.  
[https://doi.org/10.1016/S0047-2352\(02\)00132-0](https://doi.org/10.1016/S0047-2352(02)00132-0)