

# Fortalecimiento del SAT mediante aprendizaje de máquinas en la detección de violencia atípica

Juan Diego Heredia Niño  
Facultad de Economía  
Universidad de los Andes

22 de octubre de 2025

## Resumen

Este trabajo evalúa si los modelos de aprendizaje automático pueden complementar la detección de eventos atípicos de violencia vinculados al conflicto armado y, con ello, fortalecer el Sistema de Alertas Tempranas (SAT) de la Defensoría del Pueblo. Para abordar la pregunta, se construye un panel municipal trimestral (1997–2024) que reúne tres índices de violencia (IACV, IA e IGC), dieciocho predictores socio-económicos y geoespaciales, y, próximamente, la serie histórica de alertas del SAT. Se estiman dos algoritmos, Lasso y Bosques Aleatorios, mediante validación cruzada; sus probabilidades se incorporan como un insumo cuantitativo que complementa, sin sustituir, la evaluación cualitativa del SAT. El Bosques Aleatorios eleva el área bajo la curva ROC de 0,81 a 0,90, incrementa la precisión global de 0,72 a 0,85 y reduce la razón de falsas alarmas de 0,69 a 0,01 comparado con el Lasso. Estos resultados preliminares indican que metodologías de aprendizaje automático con las alertas existentes puede apoyar de forma sustantiva la priorización territorial de las intervenciones. El estudio ofrece un marco reproducible para anticipar eventos de violencia atípica y evidencia que la integración de aprendizaje automático con el SAT puede optimizar la priorización territorial.

**Clasificación JEL:** C53; D74; O54

**Palabras clave:** aprendizaje automático; alertas tempranas; violencia atípica; conflicto armado; Colombia

# 1. Introducción

La violencia en Colombia representa un desafío para el desarrollo económico y social, con costos que pueden superar el 3 % del PIB anual, recursos que podrían destinarse a sectores productivos y políticas sociales (BID, 2017). Además, estudios como el de Londoño and Guerrero (1999) indican que al considerar efectos indirectos, como pérdidas en productividad, deterioro del capital humano y disminución de la inversión extranjera, los costos totales pueden llegar al 7.1 % del PIB en América Latina y el Caribe.

En este contexto, identificar picos de violencia, o eventos de violencia atípica, se ha vuelto una prioridad para las instituciones responsables de proteger los derechos humanos y mantener la estabilidad social en Colombia. A diferencia de los incidentes violentos comunes, los episodios de violencia atípica se presentan de forma repentina, lo que dificulta una respuesta rápida y adecuada por parte de las autoridades y de la sociedad. Debido a su naturaleza inesperado e intenso, este tipo de violencia tiene un fuerte impacto, desde el desplazamiento forzado y la pérdida de vidas hasta la desorganización social, económica y política en las comunidades afectadas.

El Sistema de Alertas Tempranas (SAT) de la Defensoría del Pueblo es un instrumento institucional que emite advertencias sobre situaciones de vulnerabilidad, riesgo inminente y eventos violentos de carácter atípico. Este funciona a través del monitoreo constante de factores asociados con la violencia y el conflicto armado en el territorio, utilizando metodologías cualitativas, como el análisis de indicadores sociales, la recopilación de testimonios y la observación de campo. Con este enfoque, el SAT busca identificar dinámicas sociales y políticas que puedan favorecer el surgimiento de hechos violentos, de modo que las autoridades y los actores sociales puedan actuar de forma coordinada para prevenir o mitigar sus consecuencias.

La cobertura geográfica del SAT abarca desde áreas urbanas hasta zonas rurales remotas, lo que exige una coordinación constante con administraciones locales, la fuerza pública y organizaciones de la sociedad civil. Garantizar la fiabilidad de la información y su correcta interpretación en contextos diversos es un desafío considerable. Por ello, el SAT se sustenta en la recolección directa de datos, el seguimiento de fuentes locales y la verificación de testimonios, procesos que requieren una intervención humana significativa y experiencia en terreno, lo cuál puede dejar por fuera del estudio variables no observables utilizando estas metodologías.

Así, la pregunta de investigación es: Pueden los modelos de aprendizaje automático mejorar la detección de eventos atípicos de violencia y, con ello, fortalecer el SAT? El aporte de este trabajo a la línea de estudio es doblemente metodológico y empírico. Primero, introduce tres indicadores de violencia, IACV, IA e IGC, que permiten agregar la intensidad del conflicto a lo largo del tiempo y entre los municipios del país. Segundo, incorpora dos algoritmos de aprendizaje automático, Lasso y Bosques Aleatorios, capaces de estimar de manera trimestral la probabilidad de ocurrencia de violencia atípica en cada municipio.

Los resultados preliminares muestran que el Bosques Aleatorios eleva el área bajo la curva ROC del enfoque actual de 0,74 a 0,90, incrementa la precisión global de 0,72 a 0,85 y reduce la proporción de falsas alarmas de 0,69 a 0,01, mientras que el Lasso alcanza un AUC de 0,81 con sensibilidad de 0,74 y especificidad de 0,71. La mejora proviene de la capacidad del algoritmo para explotar patrones espacio-temporales y relaciones no lineales que la metodología cualitativa no puede procesar en tiempo real.

El estudio aporta así un marco empírico para anticipar eventos de violencia atípica a nivel nacional, demuestra que la combinación de aprendizaje automático y análisis cualitativo puede incrementar de forma sustantiva la precisión del SAT y ofrece evidencia concreta para sustentar la adopción de herramientas cuantitativas en las rutinas de monitoreo de las instituciones de derechos humanos en Colombia.

## **2. Revisión de Literatura**

Los estudios que buscan anticipar la violencia ligada al conflicto armado siguen dos caminos. El primero se centra en la persistencia de la violencia donde ya existe, mientras el segundo intenta prever eventos atípicos. Este trabajo pertenece a la segunda línea y se relaciona con varias investigaciones recientes.

Los primeros esquemas estadísticos, como el Political Instability Task Force, usaron regresiones logísticas con datos país-año para anticipar guerras civiles (Goldstone et al., 2010). Aquellos modelos fueron pioneros, pero enfrentaron el problema de desbalance de muestra: la mayoría de naciones no entra en guerra cada año. Con información sub-nacional, Ward et al. (2013) y Hegre, Hultman y Nygård (2019) mejoraron un poco la precisión al usar datos mensuales, aunque advirtieron que casi toda la capacidad predictiva venía de la inercia histórico-espacial.

En Colombia, Bazzi, Blattman y Dercon (2019) compararon Lasso, bosques aleatorios, gradiente impulsado y redes neuronales con paneles municipales. Los modelos localizaron bien focos crónicos de violencia, pero fallaron al anticipar estallidos en municipios antes pacíficos. Radford (2022), usando una red ConvLSTM, halló que el aprendizaje profundo baja el error medio pero no reduce las “sorpresas” cuando la violencia cambia de patrón. En suma, los modelos basados solo en datos pasados tienden a predecir continuidad.

Para manejar este sesgo, trabajos recientes han diseñado estrategias para el fuerte desbalance de clases que caracteriza a los eventos atípicos. El sistema ViEWS, por ejemplo, añadió el conteo de protestas locales a sus modelos de conflicto en África y así mejoró la detección de focos nuevos (Rød, Hegre, Leis, 2023). Rojas y Grautoff (2023) usaron bosques aleatorios para pronosticar masacres; equilibraron la muestra con over-/undersampling y lograron un AUC de 0,88, aunque con más falsos positivos que la regresión logística. D’Orazio (2023) aplicó AutoML y, además de la predicción, reportó in-

tervalos de probabilidad para ayudar a tomar decisiones bajo incertidumbre.

En el ámbito urbano, Mohler et al. (2011) en Los Ángeles y Barreras et al. (2016) en Bogotá modelaron delitos como “procesos auto-excitantes”, donde un hecho violento aumenta la probabilidad de otro cercano. Estos modelos predicen bien zonas calientes habituales, pero no están diseñados para detectar estallidos inusuales.

El presente estudio atiende la brecha señalada por Bazzi et al. (2019) y Radford (2022). En lugar de explicar promedios históricos, se centra en eventos atípicos medidos con tres índices de violencia (IACV, IA e IGC) para 1 102 municipios entre 1997 y 2024. A diferencia de Rojas y Grautoff (2023), que se enfocan en el AUC, aquí se comparan Lasso y bosques aleatorios con validación cruzada y varias métricas a la vez (AUC = 0,90; precisión global = 0,85; falsos positivos = 0,01). El resultado muestra que es posible bajar de forma notable las falsas alarmas sin perder sensibilidad. Además, siguiendo la recomendación de D’Orazio (2023), las probabilidades generadas se integran al protocolo operativo del SAT, lo que demuestra viabilidad de uso institucional.

El diseño del panel municipal reconoce que la historia de violencia es un predictor clave (Bazzi et al., 2019), adopta el rebalanceo de clases de Rojas y Grautoff (2023) y emplea la predicción probabilística sugerida por D’Orazio (2023). Así se da crédito explícito a las innovaciones previas.

En resumen, la literatura muestra que los modelos tradicionales predicen bien la continuidad del conflicto, pero aún fallan frente a los brotes imprevistos. Este trabajo avanza en la frontera al buscar mejorar la predicción de estos eventos atípicos, reducir de forma considerable los falsos positivos y al insertar el modelo directamente en el SAT, evidenciando que el aprendizaje automático, combinado con información cualitativa y métricas adecuadas, puede mejorar la prioridad territorial de las alertas tempranas en contextos de conflicto armado.

### **3. Definciones**

Para evaluar el riesgo de violencia asociada al conflicto armado se construyen tres índices que resumen distintos aspectos del fenómeno.

#### **Índices de violencia**

El Índice Agregado de Criminalidad y Violencia (IACV) combina homicidio, extorsión, secuestro, terrorismo y masacres para cada municipio y trimestre. Cada delito se pondera con la pena promedio establecida en el Código Penal (Ley 599 de 2000), de modo que la gravedad legal determine el peso relativo.

Delito	Penal (años)	Peso (%)
Homicidio	19.0	17.04
Extorsión	11.5	10.31
Secuestro	16.0	14.35
Terrorismo	15.0	13.45
Masacres	50.0	44.84

Ponderaciones del IACV según la pena media de cada delito.

Formalmente, el IACV para el municipio  $m$  en el trimestre  $t$  se calcula como

$$IACV_{t,m} = \frac{0,17 \text{Hom}_{t,m} + 0,10 \text{Ext}_{t,m} + 0,14 \text{Sec}_{t,m} + 0,13 \text{Terr}_{t,m} + 0,44 \text{Masc}_{t,m}}{\text{Población}_{t,m}}$$

El Índice de Amedrentamiento (IA) mide hechos que generan miedo o intimidación —amenazas, tentativas de asesinato, atentados, desplazamiento forzado y hostigamiento— cuyo impacto principal es psicológico y social.

El Índice de Gobernanza Criminal (IGC) captura la capacidad de grupos armados ilegales para imponer reglas y restringir la actividad civil mediante confinamientos, retenes ilegales, paros armados y extorsiones.

## Definición de violencia atípica

Siguiendo a Bazzi et al. (2019), se define la violencia atípica como un nivel que excede el promedio y una desviación estándar del propio municipio en los doce trimestres previos:

$$\text{ViolenciaAtípica}_{t,m} = \begin{cases} 1 & \text{si } IACV_{t,m} \geq IACV_{t-1:t-12,m} + \sigma_{t-1:t-12,m}, \\ 0 & \text{en otro caso.} \end{cases}$$

Este umbral convierte la serie continua en una variable binaria que marca trimestres con picos inusuales de violencia. Dichos eventos son poco frecuentes, lo que reproduce el desbalance propio de los fenómenos que el SAT necesita anticipar y constituye la variable objetivo de los modelos predictivos.

## 4. Hechos Estilizados

A continuación, se describen algunos hechos estilizados sobre la dinámica de la violencia y la presencia de grupos armados en Colombia, derivados de información recopilada por el Ministerio de Defensa y la Fiscalía.

La figura 1 muestra un mapa de los municipios de Colombia, distinguiendo entre los que no tienen presencia de grupos armados y aquellos que cuentan con al menos uno. Se observa que, de 2020 a 2023, aumenta el número de municipios con actores armados, sobre todo en zonas alejadas de los centros urbanos y en corredores estratégicos vinculados a economías ilegales.

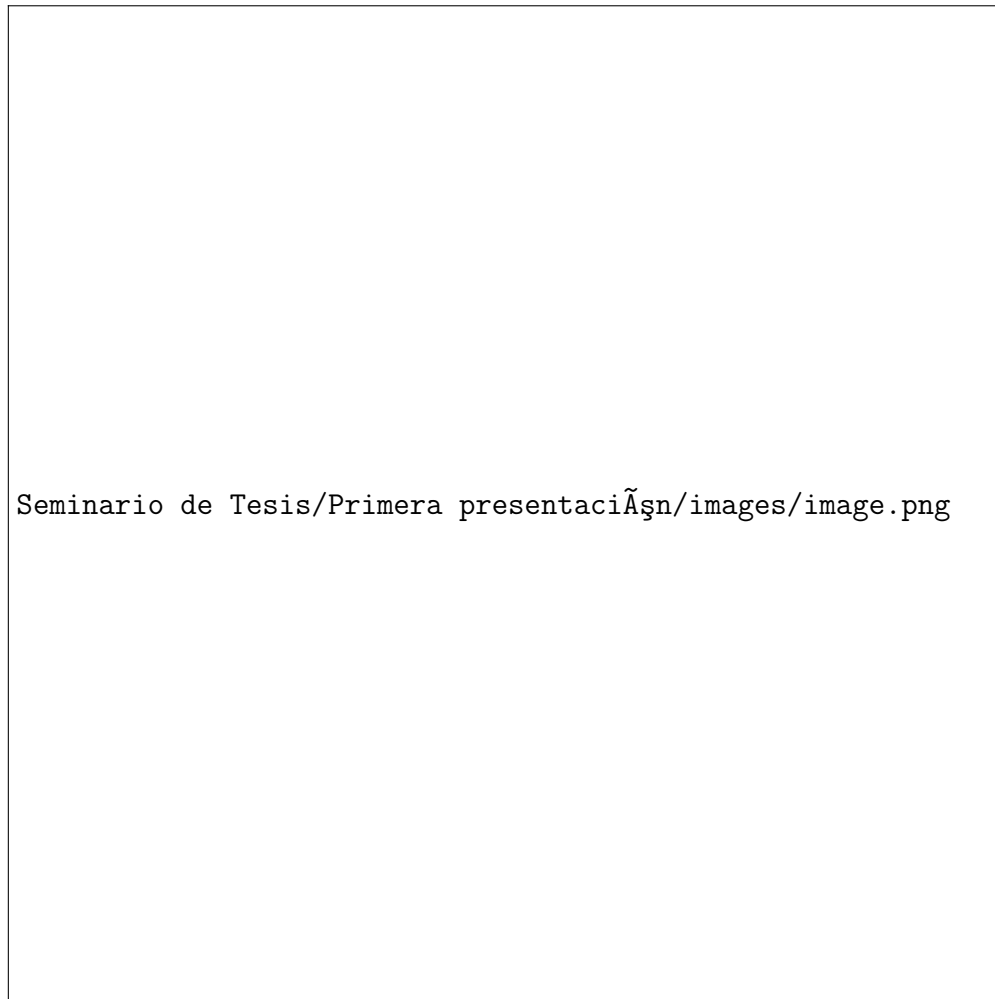


Figura 1: Clasificación binaria de los municipios según la presencia de grupos armados.

En cambio, la figura 2 emplea un degradado de color para mostrar cuántos grupos armados hay en cada municipio, desde cero hasta tres al mismo tiempo. Los tonos rojos más intensos indican lugares con varios grupos, lo que refleja una dinámica de violencia más compleja y un mayor riesgo de enfrentamientos. Así, se puede observar la expansión y consolidación de las zonas conflictivas a lo largo del tiempo, evidenciando el incremento de la superposición de grupos en distintas áreas.

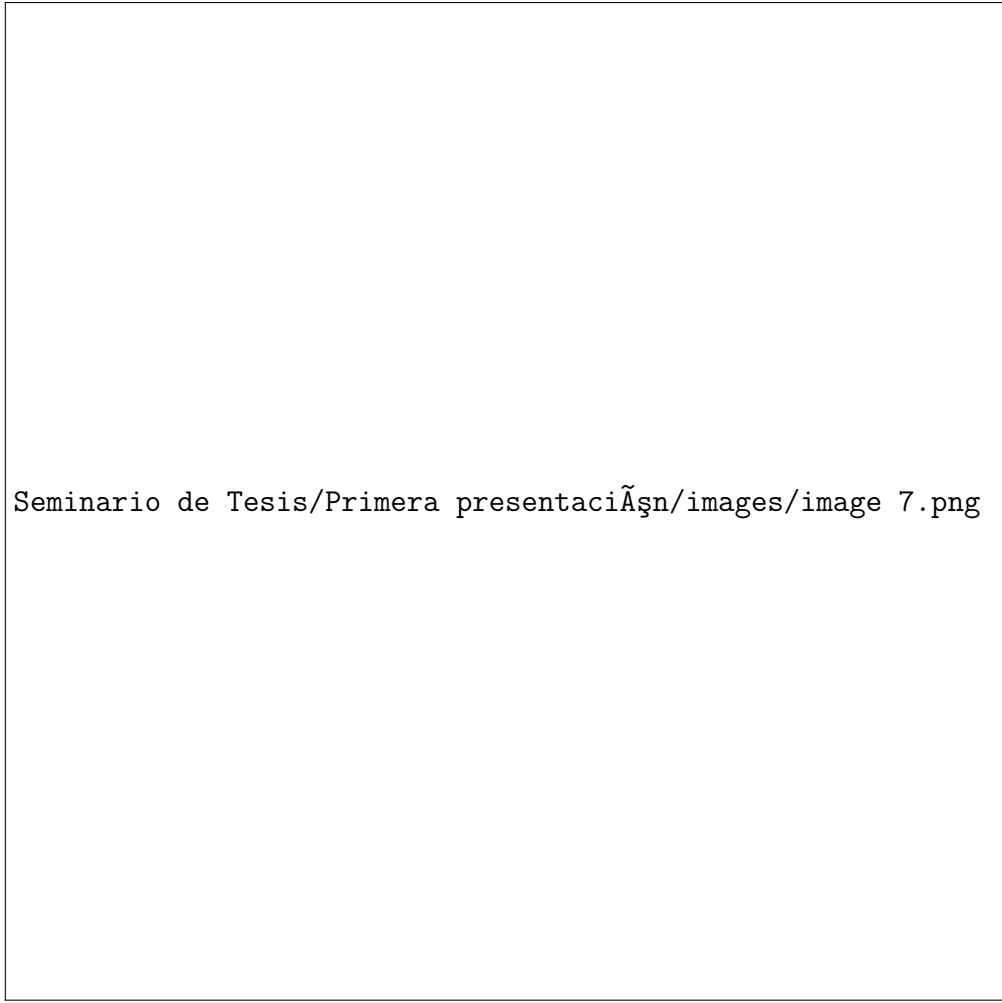
Por su parte, la figura 3 muestra la evolución del IACV a lo largo del tiempo, comparando los datos de la Fiscalía General de la Nación y del Ministerio de Defensa Nacional. Se observa una tendencia en aumento desde 2020 hasta 2024, aunque con diferencias en la magnitud y variabilidad entre las dos fuentes. Esto sugiere que la expansión de gru-



Figura 2: Concentraci3n de grupos armados por municipio.

pos armados en Colombia a trav3s de los a1os ha venido acompa1ado de un aumento agregado de la violencia en Colombia.

Asimismo, La figura 4 y 5 muestran la Curva de Lorenz del IACV para el a1o 2019 y 2024 respectivamente. Esta curva sirve para analizar c3mo se reparte la violencia entre la poblaci3n. Cuanto m3s se aleja la curva de la l3nea de igualdad perfecta (en rojo), mayor es la concentraci3n en un grupo peque1o de la poblaci3n. Para 2019, el 3ndice de Gini es de 66 %, lo que indica que la violencia se concentra en pocos municipios. En contraste, para 2024, el 3ndice de Gini sube a 73 %, reflejando que la concentraci3n de la violencia es a1n mayor. As3, en conjunto con los gr3ficos anteriores, se sugiere que este aumento de violencia, que ha sido acompa1ado de una expansi3n de grupos criminales en el pa3s, no solo ha aumentado, si no que ha venido aumentando y concentrandose en una menor cantidad de municipios.



Seminario de Tesis/Primera presentaciÃ³n/images/image 7.png


Figura 3: Evoluci3n del IACV seg3n los registros de la Fiscalía General de la Naci3n y el Ministerio de Defensa Nacional.

Para cerrar, la figura 6 y 7 muestran la relaci3n entre los quintiles de variables asociadas al desarrollo y bienestar municipal y el valor medio del IACV, junto con sus intervalos de confianza al 95 %.

Para la figura 6 se utiliza como variable objetivo el índice de necesidades básicas insatisfechas. Se observa un incremento progresivo en el promedio del IACV entre los primeros cuatro quintiles, alcanzando su punto más alto en el cuarto. Por su parte, la figura 7 utiliza como variable objetivo el índice de alfabetizaci3n municipal. Se observa una disminuci3n del promedio del IACV a partir del segundo quintil.

Estos resultados indican que la violencia tiende a concentrarse en municipios con niveles de carencias elevados, pero no extremos. En los municipios con las mayores privaciones, quinto quintil en la figura 6 y primer quintil en la figura 7, el IACV presenta una ligera reducci3n. Asimismo, los intervalos de confianza evidencian la variabilidad de cada grupo y sugieren diferencias estadísticamente significativas.





Seminario de Tesis/Primera presentaciÃ³n/images/image 3.png


Figura 4: Curva de Lorenz del IACV para 2019.

## 5. Datos

Los datos de violencia provienen de tres fuentes principales. La Fiscalía General de la Nación proporciona registros de homicidio, extorsión, secuestro, terrorismo y masacres a nivel municipal con frecuencia mensual desde 2014 hasta 2024. El Ministerio de Defensa Nacional también reporta estas mismas categorías de delitos, pero con una serie temporal más extensa, desde 1997 hasta 2024, permitiendo un análisis de tendencias de largo plazo.

Adicionalmente, se utilizarán datos de la Jurisdicción Especial para la Paz (JEP) desde 2017 hasta 2024 sobre la presencia de grupos armados, así como eventos relacionados con la violencia y el control territorial, incluyendo homicidios, amenazas, emboscadas, tentativas de asesinato, atentados, desplazamientos forzados, hostigamientos, confinamientos, extorsiones, retenes ilegales y paros armados. Utilizando estas variables se puede mejorar la caracterización de la violencia, ya que no solo considera delitos registrados formalmente, sino también eventos asociados a la dinámica del conflicto armado.

Para contextualizar la violencia dentro de las características estructurales de los municipios, se incorporan datos del Panel Municipal del CEDE, con información anual dispo-




Seminario de Tesis/Primera presentaci3n/images/image 4.png

Figura 5: Curva de Lorenz del IACV para 2024.

nible desde 2005 hasta 2023. Este conjunto de datos contiene variables que capturan diferentes dimensiones demográficas, socioeconómicas e institucionales. Entre ellas se incluyen medidas de pobreza multidimensional, acceso a servicios básicos, educación, condiciones de infraestructura, salud, programas de atención a víctimas del conflicto y características territoriales como la distancia a mercados y la altura sobre el nivel del mar.

Además, se incorpora información sobre cultivos ilícitos de coca desde 1999 hasta 2023, proveniente del Observatorio de Drogas de Colombia del Ministerio de Justicia. La presencia de estos cultivos ha sido relevante en la literatura como un determinante de violencia y financiación de grupos armados.

Finalmente se utiliza información de luminosidad nocturna obtenida del VIIRS Nighttime Light, con registros mensuales a nivel municipal desde 2012 hasta 2023. La luminosidad nocturna es un proxy ampliamente utilizado en estudios económicos para estimar el nivel de ingreso y desarrollo económico en regiones donde los datos son limitados.



Seminario de Tesis/Primera presentaciÃ³n/images/output1.png

Figura 6: Necesidades Básicas Insatisfechas vs. IACV con Intervalos de Confianza (95 %).

## 6. Metodología

### Aprendizaje Automático


El propósito es estimar, para cada municipio  $m$  y trimestre  $t$ , la probabilidad de que ocurra un evento de violencia atípica. Formalmente se busca la función

$$\hat{P}(\text{ViolenciaAtípica}_{t,m} = 1 \mid \mathbf{X}_{t,m}),$$

donde  $\mathbf{X}_{t,m}$  agrupa predictores estructurales y rezagos de violencia.

### Clasificación versus regresión

Se emplea un enfoque de clasificación binaria porque la variable dependiente toma solo dos valores (violencia atípica: sí o no). Los clasificadores permiten estimar probabilidades que se interpretan como riesgos y facilitan la fijación de umbrales operativos. Estos umbrales, a su vez, pueden ajustarse para privilegiar la sensibilidad o la precisión de acuerdo con la capacidad institucional. Además, los resultados son fácilmente integrables en esquemas de priorización de alertas tempranas.



Seminario de Tesis/Primera presentaciÃ³n/images/output2.png

Figura 7: Alfabetizaci3n vs. IACV con Intervalos de Confianza (95 %).

## Especificaci3n del modelo

La probabilidad se representa de forma general como

$$\text{ViolenciaAtípica}_{t,m} = f(IACV_{t-1:t-12,m}, IGC_{t-1:t-4,m}, IA_{t-1:t-4,m}, \mathbf{S}_{t,m}),$$

donde la funci3n  $f(\cdot)$  se aproxima con dos algoritmos de aprendizaje autom3tico.

## Modelos a utilizar

El primero es Lasso, una regresi3n logística con penalizaci3n  $L_1$  que selecciona de manera autom3tica los predictores lineales m3s relevantes (Tibshirani, 1996). Estos modelos lineales mejoran la interpretabilidad del modelo. Asί, Lasso simplifica la selecci3n de variables sin comprometer la precisi3n de las predicciones. Adem3s, es 3til para datos con alta dimensionalidad, evitando el sobreajuste mediante la regularizaci3n de los coeficientes. Sin embargo, su principal limitaci3n es que s3lo captura relaciones lineales entre las variables, lo que podría afectar su desempeño en la predicci3n de eventos atípicos si existen interacciones no lineales complejas. Tambi3n requiere una adecuada calibraci3n de sus hiperpar3metros, lo que aumenta el costo computacional del entrenamiento.

Bosques Aleatorios, propuesto por Breiman (2001), es un modelo basado en la combinación de múltiples árboles de decisión, lo que mejora la robustez y precisión del modelo en comparación con un árbol único.

Así, es capaz de manejar datos con relaciones no lineales y capturar interacciones complejas entre variables. Además, es resistente al ruido y a la presencia de valores faltantes. Sin embargo, su desventaja principal es la falta de interpretabilidad, ya que la combinación de múltiples árboles dificulta la explicación del proceso de toma de decisiones del modelo. Además, puede ser computacionalmente costoso en grandes volúmenes de datos.

## Procedimiento de estimación

La base de datos se divide en cinco pliegues y se realiza validación cruzada temporal. Dada la naturaleza atípica de estos eventos de violencia, el desbalance se corrige mediante sobremuestreo de la clase minoritaria en el conjunto de entrenamiento. Los hiperparámetros de cada algoritmo se ajustan con validación cruzada, optimizando el F1Score. Para evaluar el desempeño se reportan AUC, precisión global, sensibilidad, precisión y tasa de falsos positivos.

La probabilidad estimada  $\hat{P}(\text{ViolenciaAtípica}_{t,m} = 1)$  se incorpora como puntaje de riesgo en el protocolo del SAT. Los umbrales de decisión se ajustan según las prioridades de política: reducir falsos negativos cuando se teme subestimar eventos críticos o reducir falsos positivos para evitar alertas innecesarias que puedan sobrecargar la capacidad de intervención.

## Comparación de Modelos y Evaluación de Desempeño

En el campo del aprendizaje automático, es común comparar múltiples modelos para seleccionar el más adecuado según el problema específico. Cada metodología mencionada tiene fortalezas y debilidades, por lo que su desempeño se evaluará utilizando métricas estándar como precisión, sensibilidad, F1Score y AUC.

Para evaluar el desempeño de los modelos en este estudio, se emplearán las siguientes métricas: precisión, sensibilidad, F1Score y AUC.

### Precisión: Fiabilidad en la Identificación de Eventos Atípicos

La precisión mide qué proporción de las predicciones positivas realmente corresponden a eventos violentos atípicos. Matemáticamente, se expresa como:

$$\text{Precisión} = \frac{\text{Verdaderos Positivos}}{\text{Verdaderos Positivos} + \text{Falsos Positivos}} \quad (1)$$

Desde una perspectiva operativa, una alta precisión significa que el modelo comete pocos errores al identificar eventos atípicos, minimizando el número de falsas alarmas. Esto es fundamental para evitar la sobrecarga de recursos de seguridad, asegurando que los esfuerzos preventivos se concentren en situaciones de riesgo real.

### **Sensibilidad: Capacidad de Detectar Eventos Atípicos Reales**

La sensibilidad, por otro lado, mide qué proporción de los eventos atípicos reales son correctamente detectados por el modelo. Se define como:

$$\text{Sensibilidad} = \frac{\text{Verdaderos Positivos}}{\text{Verdaderos Positivos} + \text{Falsos Negativos}} \quad (2)$$

Una alta sensibilidad es prioritaria en contextos donde el costo de no identificar un evento crítico es mayor que el de generar una falsa alarma. En el caso de la violencia, no detectar un evento atípico a tiempo puede traducirse en pérdida de vidas humanas y en la incapacidad del sistema de alertas para reaccionar adecuadamente ante situaciones de emergencia.

### **F1-Score: Equilibrio entre Precisión y Sensibilidad**

Dado que existe una disyuntiva entre precisión y sensibilidad, se empleará el F1-Score como métrica de referencia para evaluar el desempeño global del modelo. El F1-Score es la media armónica entre precisión y sensibilidad, proporcionando un balance entre ambos aspectos. Su fórmula es la siguiente:

$$F1 = 2 \times \frac{\text{Precisión} \times \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}} \quad (3)$$

Un F1-Score alto indica que el modelo es capaz de detectar la mayoría de los eventos violentos atípicos sin generar un número excesivo de falsas alarmas. En el contexto del presente estudio, esta métrica es clave, ya que permite evaluar si el uso de aprendizaje de máquinas mejora la capacidad predictiva respecto a los métodos tradicionales de alertas tempranas.

Además, el uso del F1-Score ayuda a equilibrar dos aspectos fundamentales: la eficiencia en la asignación de recursos y la necesidad de priorizar la protección de la población ante posibles episodios de violencia extrema.

### **Área bajo la curva ROC (AUC): capacidad de discriminación independiente del umbral**

Esta métrica resume, en un único valor, todas las combinaciones posibles de sensibilidad y tasa de falsos positivos que se obtienen al variar el umbral de clasificación. Al integrar la curva Receiver Operating Characteristic (ROC),

$$\text{AUC} = \int_0^1 \text{Sensibilidad}(\text{Precisión}) \, d(\text{Precisión}),$$

se mide la probabilidad de que el modelo asigne una puntuación de riesgo más alta a un municipio con violencia atípica que a uno sin ella. Por no depender de un umbral específico, la AUC permite comparar modelos en términos de su capacidad inherente para ordenar los casos según su probabilidad subyacente de un evento crítico. En contextos con fuerte desbalance de clases, como el presente, esta característica resulta valiosa porque evita ajustes arbitrarios y se centra en la calidad de la discriminación estadística. Una AUC cercana a 1 indica que el algoritmo clasifica correctamente la mayoría de los pares municipio-trimestre, mientras que un valor próximo a 0,5 sugiere desempeño no mejor que el azar.

En el marco del SAT, la AUC facilita priorizar intervenciones al ofrecer un ranking robusto de municipios según su riesgo estimado, sin necesidad de fijar de antemano el punto de corte que activaría una alerta.

## 7. Resultados preliminares

	<b>Lasso</b>	<b>Bosques Aleatorios</b>
AUC	0.811	0.898
Sensibilidad	0.743	0.604
Precisión	0.707	0.997
Precisión global	0.720	0.854
Relación FP/TP	0.687	0.009
Relación FN/TP	0.345	0.656

Cuadro 1: Desempeño comparado de los modelos (validación cruzada temporal).

La Tabla ?? presenta los indicadores de desempeño para los dos algoritmos estimados. El modelo Lasso arroja un área bajo la curva ROC (AUC) de 0.81 y combina una sensibilidad de 0.74 con una precisión de 0.71, lo que se traduce en una precisión global de 0.72. Estos valores sugieren que, bajo relaciones lineales, el conjunto de predictores históricos y estructurales ofrece una capacidad moderada para identificar brotes de violencia atípica.

El Bosques Aleatorios, por su parte, eleva la AUC a 0.90 y la precisión global a 0.85. La precisión alcanza 0.997, lo cual reduce de forma drástica la proporción de falsas alarmas ( $FP/TP = 0.009$ ). Sin embargo, la sensibilidad desciende a 0.60 y, en consecuencia, la relación de falsos negativos aumenta. En términos operativos, el modelo basado en árboles discrimina mejor los municipios verdaderamente críticos, aunque pasa por alto algunos eventos menos evidentes.

La ganancia en precisión refleja la capacidad del Bosques Aleatorios para capturar interacciones no lineales entre los índices de violencia, los factores estructurales y la dinámica reciente del conflicto. La menor sensibilidad sugiere que, si la prioridad institucional

es evitar omisiones, podría ser necesario ajustar el umbral de decisión o incorporar técnicas complementarias que eleven la detección de casos positivos.

En síntesis, los resultados preliminares indican que el Bosques Aleatorios supera al Lasso en la mayoría de las métricas relevantes, especialmente en la reducción de alertas erróneas. Esta mejora ofrece un argumento sólido para integrar el puntaje de riesgo del modelo en el protocolo del SAT, siempre que se calibren los umbrales de alerta de acuerdo con los objetivos de política pública y la capacidad de respuesta disponible.

## **8. Conclusiones y líneas de trabajo futuro**

La evidencia preliminar muestra que los modelos de aprendizaje automático pueden complementar el Sistema de Alertas Tempranas (SAT) al estimar con mayor precisión la probabilidad de violencia atípica a nivel municipal. Entre los dos algoritmos evaluados, el Bosques Aleatorios destaca por un área bajo la curva ROC de 0.90 y una reducción de 98 % en la proporción de falsas alarmas respecto al Lasso. Estos resultados sugieren que estimaciones no lineales y pronósticos permiten identificar con mayor certeza los municipios más expuestos a eventos atípicos, lo cual puede optimizar la asignación de recursos preventivos.

Al mismo tiempo, la menor sensibilidad del Bosques Aleatorios advierte sobre el riesgo de omitir algunos eventos. Otro desafío consiste en traducir las probabilidades en protocolos claros para el personal del SAT, de modo que la información sea accionable y se evite la sobrecarga de alertas.

Las líneas de trabajo futuro se orientan en tres direcciones. En primer lugar, se planea evaluar algoritmos adicionales como XGBoost y LSTM para contrastar la robustez de los hallazgos y, en particular, mejorar la sensibilidad sin sacrificar especificidad. En segundo lugar, se propone incorporar fuentes de datos textuales, provenientes de los informes del SAT, para poder complementar las señales tempranas sobre tensiones locales.

Finalmente, este estudio aporta un marco reproducible que combina aprendizaje automático y análisis cualitativo para fortalecer las alertas tempranas de violencia; las extensiones previstas buscan consolidar su utilidad práctica y ampliar la capacidad de la Defensoría del Pueblo para proteger a las poblaciones en riesgo.



## Referencias

- Banco Interamericano de Desarrollo. (2017). *Informe sobre el costo de la criminalidad en América Latina*. Banco Interamericano de Desarrollo.
- Barreras, F., Díaz, C., Riascos, Á. J., & Ribero, M. (2016). *Una comparación de diferentes modelos para la predicción del crimen en Bogotá*. Universidad de los Andes, CESED.
- Bazzi, S., Blair, R. A., Blattman, C., Dube, O., Gudgeon, M., & Peck, R. (2019). *The Promise and Pitfalls of Conflict Prediction: Evidence from Colombia and Indonesia* (NBER Working Paper 25980). National Bureau of Economic Research.
- D'Orazio, V. (2023). *Probabilistic Conflict Forecasting with Automated Machine Learning*. Contribution to ViEWS Prediction Challenge 2023/24, West Virginia University.
- Feldmann, A. E., & Hinojosa, V. M. (2009). Terrorism in Colombia: Logic and Sources of a Multidimensional and Ubiquitous Phenomenon. *Terrorism and Political Violence*, 21(1), 42–61.
- Giraldo Alegría, S. A., Ordoñez Palacios, L. E., Bucheli, V., & Ordoñez, H. (2020). Modelo de redes neuronales para predecir la tendencia de víctimas de secuestro en Colombia. *Investigación e Innovación en Ingenierías*, 8(3), 38–49. <https://doi.org/10.17081/invinno.8.3.4702>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Levitt, S., & Rubio, M. (2000). *Understanding Crime in Colombia and What Can Be Done About It*. Documento de trabajo, Universidad de Chicago/Fedesarrollo.
- Mohler, G., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. (2011). Self-Exciting Point Process Modeling of Crime. *Proceedings of the National Academy of Sciences*, 108(19), 7743–7748. <https://doi.org/10.1073/pnas.1015346108>
- Musumba, M., Fatema, N., & Kibriya, S. (2021). Prevention Is Better Than Cure: Machine Learning Approach to Conflict Prediction in Sub-Saharan Africa. *Sustainability*, 13(13), 7366. <https://doi.org/10.3390/su13137366>
- Ordoñez Eraso, H. A., Pardo Calvache, C. J., & Cobos Lozada, C. A. (2020). Detection of Homicide Trends in Colombia Using Machine Learning. *Revista Facultad de Ingeniería (University of Antioquia)*, 29(54), e105.
- Radford, B. J. (2022). High resolution conflict forecasting with spatial convolutions and long short-term memory. *International Interactions*, 48(4), 739–758. <https://doi.org/10.1080/03050629.2022.2031182>

- Riascos, Á. J., & Vargas, J. F. (2011). *Violence and Growth in Colombia: A Review of the Quantitative Literature* (Documentos de Trabajo No. 8806). Universidad del Rosario.
- Rod, E. G., Hegre, H., & Leis, M. R. (2023). Predicting armed conflict using protest data. *Journal of Peace Research*. <https://doi.org/10.1177/00223433231186452>
- Rojas, J., & Grautoff, C. (2023). *Predicción de las masacres en Colombia empleando inteligencia artificial*. Universidad de los Andes, Working Paper.
- Rubio, M. (1999). *Crimen e impunidad: Precisiones sobre la violencia*. Documento CEDE, Universidad de los Andes.
- Rubio, M. (2000). *La Violencia en Colombia: Precisiones económicas*. Inter-American Development Bank.
- Rubio, M. (2003). *El rapto de la pesca milagrosa: breve historia del secuestro en Colombia* (Documento de trabajo No. 2003-36). CEDE, Universidad de los Andes.
- Sánchez, F., & Núñez, J. (2001). *Determinantes del crimen violento en un país altamente violento: El caso de Colombia*. CEDE, Universidad de los Andes.
- Sánchez, F., & Chacón, M. (2005). Conflicto, Estado y descentralización: Del progreso social a la disputa armada por el control local, 1974–2002. *Documentos CIDER, Universidad de los Andes/LSE*.
- Urcuqui, C., Moreno, J. S., Montenegro, C., Riascos, Á., & Dulce, M. (2020). Accuracy and Fairness in a Conditional Generative Adversarial Model of Crime Prediction. *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 1631–1636. <https://doi.org/10.1109/SMC42975.2020.9283477>
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).