


Comprendiendo la dinámica de los conflictos en américa latina: una aproximación desde el machine learning

Understanding the dynamics of conflicts in latin america: an approach from machine learning

Juan José Villar Roldán ¹; Juan Manuel Martín-Álvarez ²

¹ Universidad Internacional de La Rioja, España (juanjose.villar749@comunidadunir.net) 

² Universidad Internacional de La Rioja, España (juanmanuel.martin@unir.net) 

RESUMEN

El presente estudio tiene como finalidad la identificación de patrones en los conflictos de América Latina desde el año 1989 hasta la actualidad. El artículo se basa en el supuesto de que la clusterización puede utilizarse para alcanzar una mayor comprensión sistémica de las correlaciones entre política, economía y conflicto. Se parte del supuesto de que las variables utilizadas forman parte de un sistema interactivo con correlaciones aún por comprender. La clusterización es la herramienta que se utiliza para clasificar entidades en grupos con el fin de buscar explicaciones basadas en características transversales a los objetos en los que están integradas; de este modo, el análisis busca una explicación más tangible de los complejos vínculos entre las variables de desarrollo humano, económicas y las relacionadas con los conflictos. En el análisis se utilizan los datos del Uppsala Conflict Dataset Program para categorizar

Recibido: 25 de mayo del 2023. Aceptado: 26 de mayo del 2023. Publicado: 31 de julio 2023

Este es un artículo de acceso abierto, distribuido bajo los términos de la licencia Creative Commons Atribución - No Comercia_ Compartir Igual 4.0 Internacional. (<http://creativecommons.org/licenses/by-nc-sa/4.0/>) que permite el uso no comercial, distribución y reproducción en cualquier medio, siempre que la obra original sea debidamente citada

Juan José Villar Roldán

Licenciado en Ingeniería Industrial por la Universidad Carlos III de Madrid (UC3M) desarrolló su Tesis de Licenciatura en análisis de series temporales de indicadores macroeconómicos y su predicción. Máster en Business Intelligence por la Universidad Internacional de La Rioja (UNIR) que cuenta con amplia experiencia en consultoría de datos para los sectores bancario y energético en Londres y España.

Juan Manuel Martín-Álvarez

Doctor en Economía con enfoque en análisis cuantitativo para la toma de decisiones. Cuenta con amplia experiencia como docente en universidades públicas y privadas en las áreas de Contabilidad, Finanzas, Estadística y Econometría. Es Profesor Titular de Métodos Cuantitativos para la Economía y la Empresa en la Universidad Internacional de La Rioja (UNIR). Actualmente, es director del Máster en Business Intelligence de la Universidad Internacional de La Rioja (UNIR). Finalmente, ha publicado más de 20 artículos de Economía Aplicada, con especial enfoque en el consumo de tabaco, en revistas indexadas.

actores presentes en conflictos en función de una serie de características.

Palabra clave: Análisis clúster, aprendizaje automático, estudios de guerra, análisis económico.

ABSTRACT

The purpose of this document is to identify patterns in conflicts in Latin America from 1989 to the present. The article assumes that clustering can be used to achieve a greater systemic understanding of the correlations between politics, economics, and conflict. It starts from the assumption that the variables used are part of an interactive system with correlations yet to be understood. Clustering is the tool used to classify entities into groups to search for explanations based on cross-sectional characteristics of the objects in which they are integrated; thus, the analysis seeks a more tangible explanation of the complex links between economic, human development and conflict-related variables. Data from the Uppsala Conflict Dataset Program are used in the analysis to categorise actors present in conflicts based on a series of characteristics.

Keywords: Cluster analysis, machine learning, war studies, economic analysis.

I. INTRODUCCIÓN

En las últimas décadas, América Latina ha sido testigo de muchos conflictos que han planteado desafíos significativos para la estabilidad y el desarrollo en la región. Desde guerras civiles y enfrentamientos armados hasta la lucha contra el narcotráfico, estos conflictos han dejado una huella profunda en la región. Su comprensión y la búsqueda de enfoques efectivos para su gestión se han vuelto imperativos para promover la paz y el progreso en la zona.

Para la comprensión de dichos conflictos existen aproximaciones a las relaciones internacionales como el realismo, el liberalismo y el constructivismo que proporcionan marcos conceptuales para examinar sus causas y sus dinámicas. Desde la perspectiva realista, se podrían analizar los intereses y la competencia por el poder entre actores estatales y no estatales en la región. Las rivalidades territoriales, el control del negocio del narcotráfico y la influencia geopolítica desempeñan también un papel importante en estos conflictos.

Otra aproximación a la comprensión de la región es desde el liberalismo, que, por su parte, enfatiza la importancia de las instituciones internacionales, la cooperación y el libre comercio para prevenir y resolver los conflictos. En el contexto latinoamericano, se han observado los esfuerzos de integración regional, como la Comunidad de Estados Latinoamericanos y Caribeños (CELAC) o la Unión de Naciones Suramericanas (UNASUR), como mecanismos para fomentar la paz y la estabilidad en la región.

Por último, desde el enfoque constructivista, se puede explorar cómo las identidades, las normas y las percepciones moldean los conflictos en América Latina. Los conflictos étnicos, culturales y religiosos pueden ser analizados a través de esta perspectiva, centrándose en cómo los discursos influyen en la formación y escalada de los conflictos en la región.

Sin embargo, y a raíz de la mayor disponibilidad de datos para el análisis de los conflictos, como complemento a estas aproximaciones tradicionales de las ciencias

sociales, se hace evidente la importancia de los métodos estadísticos como complemento a las perspectivas teóricas.

Estos métodos desempeñan un papel cada vez más relevante en el análisis y la predicción de conflictos. Permiten analizar datos y tendencias para identificar patrones y factores de riesgo, brindando así una herramienta valiosa para la su comprensión y prevención. Su creciente papel invita a integrar los enfoques teóricos y empíricos para un análisis más completo.

El uso de los métodos estadísticos para el estudio de las ciencias sociales ha ido creciendo conforme el acceso a las mismas se ha ido democratizando. Una creciente bibliografía, cuyo objetivo es la predicción de la ocurrencia de eventos tales como guerras civiles (Blair, R. A. & Sambanis, N., 2020; Hegre et al., 2013; Hegre et al., 2021; Joshi, 2020; Qinn et al., 2023), disputas interestatales (Chen et al., 2022; Gleditsch & Ward, 2013; Lin et al., 2021; Williford & Atkinson, 2020) e inestabilidad política (Bowlsby et al., 2020; Goldstone et al., 2010; Khokhlov & Korotayev, 2022) se ha desarrollado recientemente.

Sin embargo, dichas fuentes se encuentran orientadas mayoritariamente al uso de modelos de regresión logística, en los que la variable dependiente utilizada es la existencia de conflictos y una serie de variables neokantianas independientes tales como el nivel democrático de los países o de control territorial como la cercanía entre los países, han adquirido un estatus casi paradigmático en el estudio sistemático de las relaciones internacionales (Hyde & Saudnders, 2020; Russett & Oneal, 2001).

En este contexto, la utilización de técnicas no supervisadas de aprendizaje automático como la clusterización emergen como herramientas valiosas en el análisis de conflictos en América Latina.

La clusterización ofrece una perspectiva más completa y detallada sobre los conflictos y sus implicaciones. Al examinar eventos y actores enriquecidos con información contextual de los mismos y agruparlos en clusters, se pueden identificar patrones, tendencias y factores comunes que subyacen a los conflictos en la región.

Dicha clusterización no solo permite describir y clasificar los conflictos, sino también descubrir relaciones ocultas y dinámicas complejas presentes en los datos. Al analizar conflictos históricos en América Latina, entre los que se podrían destacar el conflicto armado en Colombia, las luchas entre cárteles mexicanos o la guerra entre el gobierno peruano y la organización terrorista Sendero Luminoso, la clusterización puede revelar diferentes facetas y aspectos interrelacionados.

Así, puede ayudar a identificar grupos de eventos con características similares, como duración del conflicto, tipos de violencia, participantes involucrados, ubicación geográfica y nivel de desarrollo humano de la zona. Al agrupar eventos violentos relacionados se pueden identificar patrones comunes y factores subyacentes que contribuyen a la expansión y persistencia de este problema. Esto puede ser crucial para desarrollar estrategias de lucha contra estos actores más efectivas y coordinadas en toda la región.

En resumen, la clusterización se presenta como una herramienta prometedora para analizar los conflictos en América Latina, proporcionando una comprensión más profunda de sus dinámicas y relaciones. Desde conflictos armados prolongados hasta desafíos relacionados con la violencia de bandas con una estructura organizativa formal, la clusterización puede ayudar a aflorar los patrones y factores que impulsan estos conflictos y, en última instancia, contribuir a la formulación de políticas y estrategias más efectivas para su gestión en la región.

II. METODOLOGÍA

La principal fuente de datos utilizada para este documento es el Conjunto de Eventos Georeferenciados de Uppsala (GED, por sus siglas en inglés), compilado por el Programa de Datos de Conflictos de Uppsala (UCDP) (Sundberg & Melander, 2013). Este conjunto de datos tiene la ventaja de proporcionar información exhaustiva sobre las muertes relacionadas con conflictos a nivel geográficamente desagregado y ha sido utilizado en multitud de trabajos previos sobre el análisis de conflictos (Buhaug & von Uexkull, 2021; Duursma, 2022; Ge et al., 2022; Pettersson et al., 2021).

Para la correcta comprensión de los datos analizados, resulta esencial aclarar que la base de datos GED considera conflictos aquellos enfrentamientos armados, llamados eventos, entre dos grupos que resultan en un mínimo de 25 muertes en al menos uno de los años de la serie.

A. Obtención de los datos

Con el fin de recopilar información sobre conflictos se ha utilizado la base de datos de la UCDP. La UCDP utiliza métodos que incluyen la extracción de información de fuentes abiertas, como redes sociales y medios de comunicación, así como encuestas y cuestionarios específicos. El objetivo de esta recopilación de datos es obtener información precisa y actualizada sobre los conflictos en todo el mundo, lo que permite un análisis y comprensión más profunda de los mismos.

La información recopilada de la base de datos de la UCDP incluye detalles sobre la ubicación geográfica de los eventos, las partes involucradas, las fechas, la magnitud de las bajas y otros factores relevantes. Esta información se obtiene de fuentes confiables y se actualiza regularmente para garantizar su exactitud.

Además del GED, se han utilizado diversas fuentes de información para complementar y enriquecer la información de cada uno de los eventos:

- Cities and Armed Conflict Events (CACE) aporta información acerca de la tipología de localización para cada evento y, en el caso de las ciudades, clasifica en función del tamaño de la ciudad y su capitalidad.
- World Data indicators (UN) obteniendo las poblaciones por año y país, de forma que se pueda conocer el peso proporcional de las bajas de los eventos sobre el total de la población.
- UCDP Non-State Conflict Dataset versión 22.1 complementa la información presente en el GED categorizando el nivel de organización de los grupos no estatales implicados en los eventos
- United Nations Data Program - Human Development Reports - Human Development Index (HDI) incluye un conjunto de indicadores que informan el

nivel de desarrollo de los países. Para el análisis se ha utilizado el HDI, que mide, de forma simplificada, el nivel de desarrollo humano incorporando la esperanza de vida, la media de años de educación por persona y el GNI per cápita.

Esta información adicional aporta el contexto sobre la situación en la que se ha generado cada uno de los eventos y sobre las tipologías de los grupos implicados en los eventos. Esto permite segmentar los eventos en base a sus distintas tipologías.

La fuente de datos principal (GED) contiene información desde el año 1989 hasta el 2022. Para el resto de las fuentes se han tomado periodos temporales iguales al del GED.

B. Preparación de los datos

Con anterioridad a la realización del análisis se ha determinado qué información podía aportar mayor valor. Para ello se tomaron las distintas bases de datos y se seleccionaron los campos que eran de interés, para posteriormente desarrollar las variables que se han creado. A continuación, se realizó un proceso de data-cleaning para eliminar aquellos outliers que, por su excepcionalidad, podían descompensar el modelo. Este proceso y análisis posterior se llevan a cabo en R (R Core Team), un lenguaje orientado a la computación estadística, y en su entorno de desarrollo integrado RStudio (RStudio Team). Además, las librerías de tidyverse y cluster (Maechler et al., 2022; Wickham et al., 2019) han sido utilizadas para limpiar los datos, aplicar el algoritmo de clúster y hacer las representaciones gráficas.

Las variables seleccionadas y generadas son las siguientes:

- Duración del conflicto (DC) valores de duración de conflicto en años. Para su visualización posterior se organizan por conflicto y día agrupados en 3 grupos. Largos (más de 10 años), medios (entre 1 y 10 años) o cortos (menos de 1 año).

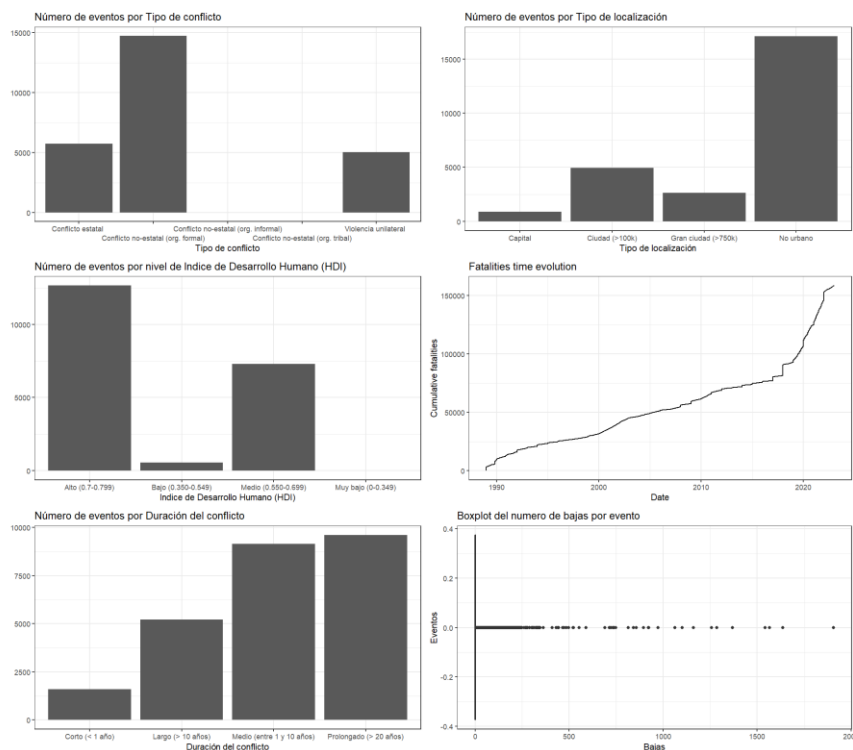
- Tipo de violencia (TV) variable compuesta utilizando los tres valores del campo tipo de violencia (one-sided violence, non-state conflict, state-based conflict) y, para los no estatales, el nivel de estructura de éstos (formal, informal, tribal, otros).
- Tipo de localización (TL) utilizando 4 valores (Capital, City (>100k), Major city (>750k), Non urban), incluyendo la categoría “No urbano” tanto eventos en zonas rurales como en núcleos urbanos de menos de 100k habitantes.
- Latitud absoluta (LA) toma los valores de latitud de los eventos en valor absoluto medir la distancia al Ecuador
- Human Development Index (HD) categorizado para su visualización en 5 grupos en función del nivel de desarrollo (Muy alto (0.8-1.0), High (0.7-0.799), Medium (0.550-0.699), Low (0.350-0.549), Very low (0-0.349)) siguiendo la metodología PNUD.
- Bajas per capita (BC) generando la variable dividiendo el nivel de bajas de cada evento entre la población del país para el año de ese evento.
- Bajas (B) el total de bajas que se han dado en el evento

Las características de los datos obtenidos tras haber generado las variables necesarias para el análisis son las siguientes:

Se observa que la mayoría de los conflictos en Latinoamérica son prolongados, con una duración mayor a 10 años (58%), frente a conflictos de duración media entre 1 y 10 años (35.80%) y conflictos cortos de menos de un año (6.24%). En cuanto al tipo de violencia se identifica un 22.5% de eventos parte de conflictos estatales, 19.7% de violencia unilateral, 57.6% de conflictos no estatales en los que los participantes tienen una organización formal, un 0.12% son conflictos no estatales con organización informal y un 0.11% con organización tribal. El número de bajas por evento está entre 1 y 1906, con una media de 6.2 y una mediana de 1. En cuanto a la localización se dan eventos en Capitales (3.45%), grandes ciudades con más de 750k habitantes (10.3%), ciudades con más de 100k habitantes (19.3%) y el resto de las zonas rurales y ciudades menores (66.9%). En la evolución temporal de las bajas se observan incrementos en la tendencia de bajas acumuladas entre el inicio de los años 90, con

un incremento de 25k bajas y desde el año 2020, en el que se percibe un incremento acusado de la tendencia, con más de 40k bajas en 2 años. El índice de desarrollo humano (HDI) se concentra en Latinoamérica en Alto y Medio para los eventos analizados. La Figura 1 representa las variables descritas anteriormente.

Figura 1. Representación de las variables utilizadas.



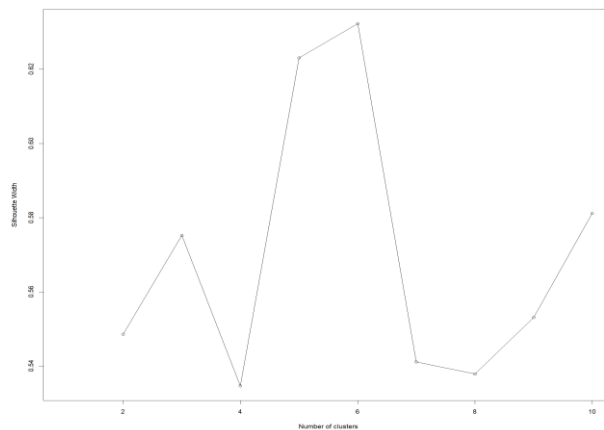
C. Clusterización de los conflictos

La clusterización es la herramienta que se utiliza para clasificar entidades en grupos con el fin de buscar explicaciones basadas en características transversales a los objetos en los que están integradas; de este modo, el análisis busca una explicación más tangible de los complejos vínculos entre las variables económicas, de desarrollo humano y las relacionadas con los conflictos. Entre los objetivos de la clusterización se encuentra la identificación de diferentes grupos que a pesar de su disparidad organizativa y geográfica se comportan de manera similar.

Para segmentar los eventos en clústeres diferenciables entre sí se utilizan técnicas de aprendizaje automático. Estas técnicas de aprendizaje no supervisado se usan para obtener información sobre los actores y eventos y pueden crear grupos de eventos con un perfil similar. Por lo tanto, utilizamos este enfoque para segmentar los eventos de conflictos en Latinoamérica en grupos.

Las técnicas de agrupación son herramientas que ayudan a comprender los diferentes subgrupos que existen dentro de un conjunto de datos. Específicamente, estas técnicas tienen como objetivo agrupar los elementos que están lo suficientemente cerca unos de otros y lo suficientemente lejos de otros elementos (Rokach & Maimon, O., 2005). Sin embargo, la elección de cómo medir la distancia entre dos elementos es algo que no está acordado en la literatura y hay muchas alternativas subjetivas. La distancia es una medida numérica que mide la distancia entre duplas de eventos, en otras palabras, mide la proximidad o similitud entre individuos.

Figura 2. Número óptimo de clústeres. Ancho de silueta.



La forma más común de medir la distancia es la distancia euclidiana, aunque existen otras alternativas como la distancia de Manhattan utilizada para tipos particulares de problemas. Sin embargo, al ser los datos utilizados mixtos (numéricos

y categóricos) estas distancias no son aplicables, por lo que se han de descartar algoritmos de agrupación tradicionales como K-means o la agrupación jerárquica.

Para nuestro estudio de caso, utilizamos la distancia de Gower, que es una medida de distancia que se puede calcular para dos individuos cuyos atributos están mezclados. La distancia de Gower se calcula como el promedio de las diferencias entre individuos. Cada distancia de Gower se encuentra entre [0, 1].

$$d(i, j) = \frac{1}{p} \sum_{f=1}^p d_{if}^{(f)}$$

La disimilitud parcial $d(i, j)(f)$ depende del tipo de variable que estemos midiendo. En el caso de las variables numéricas, la disimilitud parcial es la relación entre las diferencias absolutas entre las observaciones y el rango máximo observado de todos los individuos. En el caso de las variables categóricas, la disimilitud parcial es 1 si las observaciones son diferentes y 0 en el caso de que no lo sean.

Adicionalmente, se ha de seleccionar un algoritmo de agrupación que ha de cuadrar bien con la distancia de Gower. Para ello, seleccionamos el algoritmo k-medoids. El algoritmo k-medoid, Partitioning Around Medoids (PAM) es un método clásico de particionamiento similar al conocido método k-means pero, en lugar de iterar sobre los centroides, itera sobre los medoides, es decir, intenta encontrar el objeto más representativo para cada clúster (Rokach & Maimon, O., 2005). El algoritmo agrupa los objetos en un total de k clústeres donde k debe ser dado a priori. La selección del número óptimo de clústers (k) debe hacerse teniendo en cuenta la información estadística obtenida en los datos, aunque si existe alguna justificación razonada a priori, el número de clústers puede variar por diferentes motivos. Para seleccionar el número óptimo de clusters en los que dividir nuestros datos utilizamos el ancho de silhouette. El ancho de silhouette es una de las opciones más utilizadas para medir la similitud entre cada punto de un clúster y compara esta similitud con el

punto más cercano del cúmulo vecino. Esta métrica se encuentra entre $[-1, 1]$ donde los valores más altos significan mayores similitudes.

La Figura 2 muestra el resultado de la medición para valores de k entre 2 y 10 donde se puede observar que segmentar a los eventos en 5 y 6 grupos maximiza la similitud dentro de cada uno de los clusters y la disimilitud entre los clusters. Hemos dividido la muestra en 5 y 6 grupos siguiendo los resultados encontrados en el análisis de silueta, sin embargo, el uso de $k = 5$ produce resultados más interpretables desde un punto de vista de estudio de los conflictos.

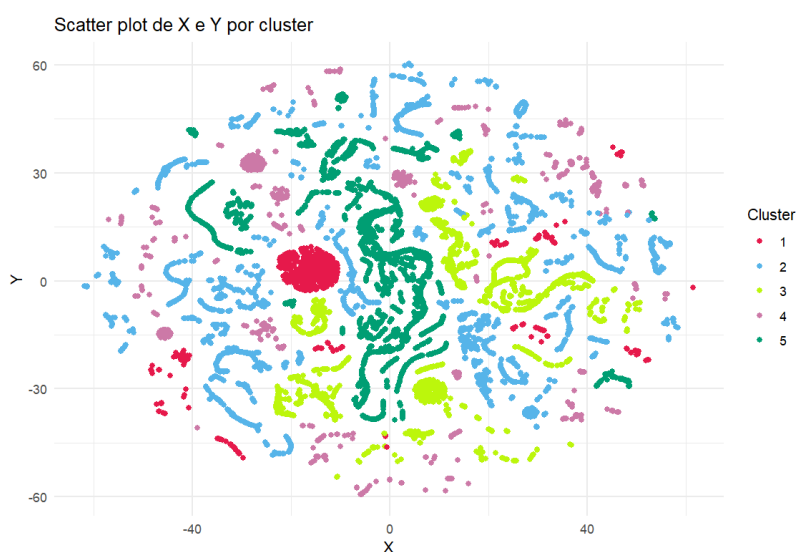
III. RESULTADOS

En esta sección se exponen los resultados obtenidos mediante la aplicación de la técnica de clustering k-medoids utilizando la distancia de Gower. Tal y como se explica en la sección de Metodología, el algoritmo de aprendizaje automático k-medoids se utiliza para agrupar datos y determinar el elemento más representativo de cada cluster. En nuestro caso de estudio, este elemento representa el evento de conflicto más recurrente dentro de cada cluster. La Tabla 1 presenta las características que definen el perfil de cada cluster utilizando k-medoids ($K=5$).

Tabla 1. Características de los k-medoids ($k=5$).

Variable	Clusterización conflictos				
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
DC	8.384	8.394	27.75	10.27	30.38
TV	No-estatal formal	No-estatal formal	Unilateral	No-estatal formal	Estatal
TL	Gran Ciudad	No urbano	No urbano	Ciudad	No urbano
LA	24.068	19.764	8.398	23.363	6.660
HD	0.7551	0.7532	0.6629	0.7539	0.6667
BC	3.7323	9.927	18.130	6.5306	46.890
B	4.202	9.154	2.939	4.501	6.677

Figura 3. Diagrama de clusters de los eventos en Latinoamérica para identificar las características de los eventos (k=5).



La Figura 3 utiliza la técnica t-Distributed Stochastic Neighbour Embedding (Rokach & Maimon, O., 2005) para visualizar la división de los eventos en cinco grupos. Esta técnica nos permite representar nuestros datos multidimensionales en un gráfico bidimensional.

Dividir la muestra en seis grupos no arroja resultados muy útiles en términos de clasificación de las tipologías de los conflictos, al no apreciarse una mejora sustancial en el valor de silhouette. Esto nos lleva a realizar el clustering con $k = 5$.

Para una mejor visualización de nuestros resultados, representamos en la figura 4, las variables en las que encontramos diferencias entre clústeres que pueden considerarse como las principales variables discriminatorias entre los 5 clústeres.

El clúster 1 aglomera a la casi totalidad de eventos sucedidos en grandes ciudades (99%), siendo la duración de los conflictos Media para un elevado porcentaje de los casos (75.1%) y superior a 10 años para un 20.06% de los eventos. Adicionalmente, la práctica totalidad de los conflictos se dan entre organizaciones no estatales con un nivel de organización formal. Los eventos presentes en el clúster son

mayoritariamente de México (75.6%) y Brasil (23.6%), llamando la atención que la totalidad de los eventos ocurridos en Argentina se sitúan en este clúster. Los eventos presentes en este clúster corresponden a violencia entre cárteles ocurrida en grandes ciudades de más de 750k habitantes, llamando la atención la incorporación de los enfrentamientos entre los Comandos Vermelhos y las fuerzas de los GED brasileños.

El clúster 2 contiene los enfrentamientos que se dan en localizaciones no urbanas en conflictos no estatales con una organización formal (98.8%). La duración del conflicto en los eventos presentes en el clúster se encuentra repartida, siendo la casi totalidad de los enfrentamientos inferiores a 20 años. Predominan los conflictos Medios con un 54.5% de los casos, seguidos de los largos con un 32.5%. Los casos de países como Paraguay (enfrentamientos entre cárteles mexicanos en territorio panameño), Bolivia (conflicto Laime – Qaqachaca) o Ecuador se encuentran en su totalidad en este clúster. Son preponderantes en cambio los casos de Brasil y México, con unas ubicaciones más alejadas del Ecuador que el anterior clúster.

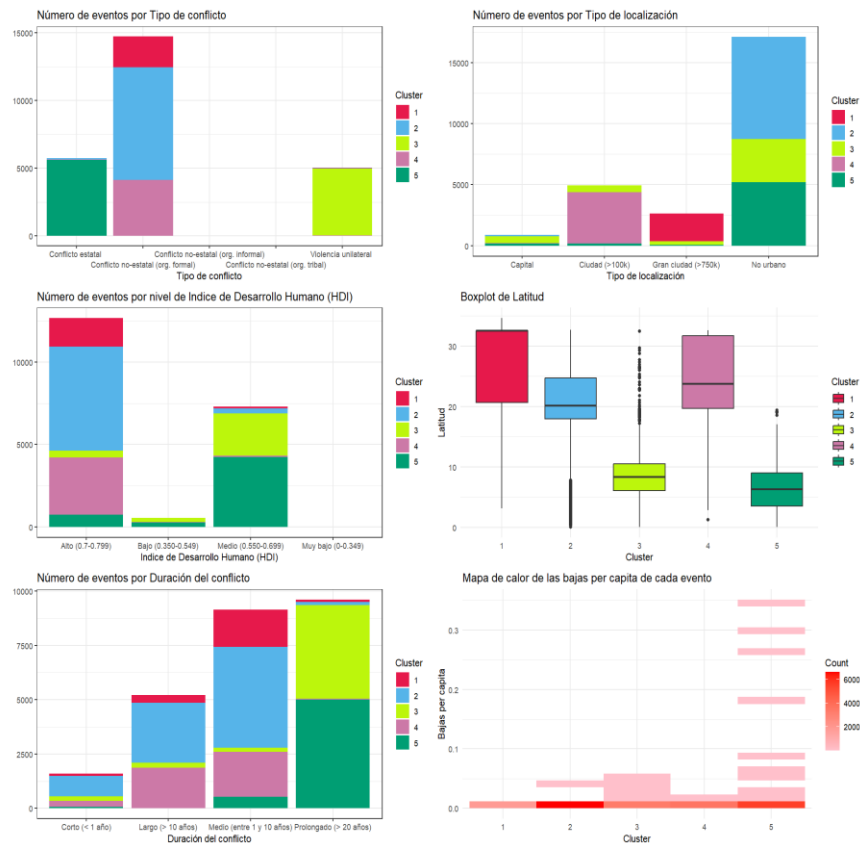
El clúster 3 comprende un porcentaje elevado de los conflictos Prolongados (52.15%) siendo la casi totalidad de los casos debidos a violencia unilateral. La predominancia de la ubicación de los eventos en zonas de HDI medio nos indica que existe una correlación entre dicho tipo de violencia y el nivel de desarrollo humano. Finalmente, se localizan principalmente en regiones no urbanas, pero cabe destacar su preponderancia en los eventos que se dan en Capitales nacionales (63.71%). Los países preponderantes en el clúster son Colombia (56.3%) y Venezuela (28.6%), destacando el hecho de que Venezuela tiene la práctica totalidad de sus eventos en el clúster (98.95%), junto con Guyana y la República Dominicana. Los eventos del clúster pertenecen eminentemente a masacres de civiles, de ahí que se pueda observar en el mapa de calor un nivel superior de bajas per cápita al de los clústeres 1 y 2.

El clúster 4 contiene eventos que se dan eminentemente en ciudades medianas de más de 100k habitantes (un 85.31% de los casos en dichas ciudades) y siendo la tipología de enfrentamiento no-estatal con grupos con organización formal. El HDI en

las ciudades en las que se han desarrollado los eventos es Alto en su mayoría (81.80%), siendo el segundo clúster que más eventos aporta al nivel Alto de HDI (27.43%). La longitud de los conflictos para este clúster se encuentra en su mayoría entre 1 y 20 años, siendo altamente improbable encontrar lo que hemos denominado un conflicto prolongado, superior a 20 años de duración (únicamente un 0.83% de los casos). En este clúster se encuentran eventos en países como Jamaica y Honduras, aunque los casos mayoritarios del clúster provienen de México y del Gobierno de Venezuela.

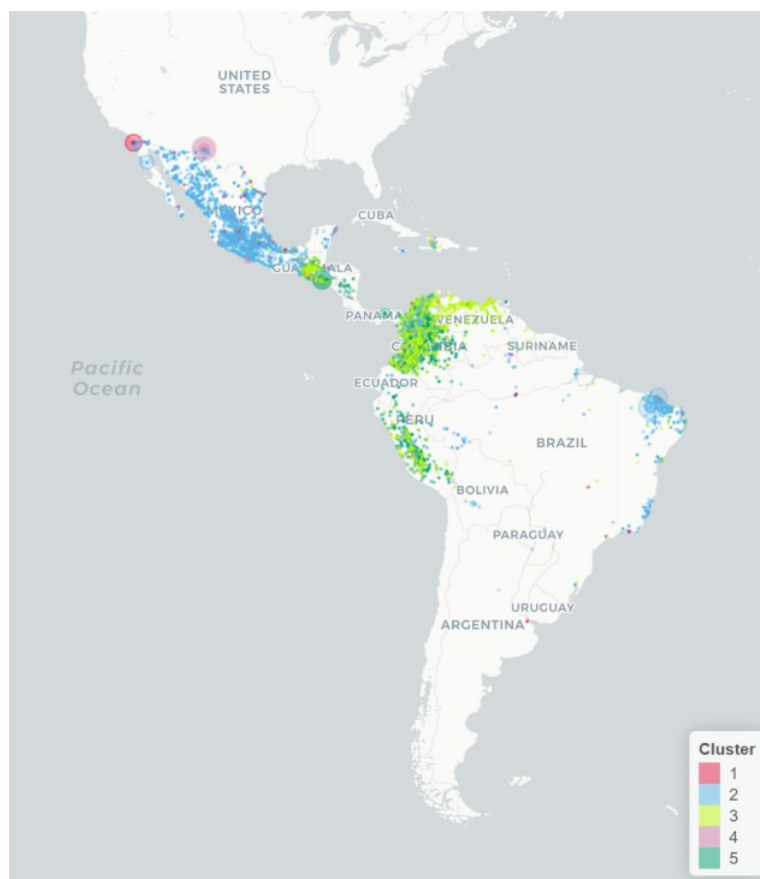
El clúster 5 agrupa los casos de conflicto estatal, que se dan mayoritariamente en regiones no urbanas (92.8%), siendo sin embargo los eventos presentes en Capitales un 20.18% del total. Se observa que los conflictos estatales son, en su mayoría, conflictos prolongados (89.1%), siendo el clúster dominante en dicha categoría de duración (52.15%). Es especialmente llamativo el nivel de bajas per cápita del clúster que, como se observa en el mapa de calor, es el que tiene los valores más extremos de los cinco. En ello influye que los conflictos mayoritarios son los enfrentamientos del gobierno de Colombia, del gobierno del Salvador y del gobierno de Guatemala.

Figura 4. Representación gráfica de la composición de cada clúster.



En resumen, el análisis nos permite identificar cinco grupos diferenciados en función de las características aportadas. Sin embargo, con el fin de comprender la ubicación de los eventos y la dispersión geográfica de los mismos para los clústeres identificados se aporta el siguiente mapa en la Figura 5.

Figura 5. Representación geoespacial de los eventos clusterizados.



En el mapa se representan los eventos geolocalizados utilizando la latitud y la longitud presentes en los datos de la base de datos GED. Esto nos permite observar los clústeres en los países en los que están presentes. El tamaño de cada uno de los puntos indica el volumen de bajas del evento. Sin entrar en análisis más pormenorizados, se observa que en la frontera con los Estados Unidos hay eventos con gran magnitud de bajas (Tijuana-San Diego), al igual que en la región de Ceará de Brasil.

De esta manera, se confirma lo observado anteriormente: la dispersión de los casos del clúster 1 por toda la geografía latinoamericana en los grandes núcleos poblacionales, la amalgama de los eventos estatales y de violencia unilateral en el

noroeste de Sudamérica (clústeres 3 y 5) y la presencia casi constante de los eventos del clúster 2 en las regiones del este de Brasil y a lo largo de México.

La presencia de los eventos a lo largo de todo el territorio mexicano y colombiano nos permiten identificar de manera visual que los conflictos existentes en dichos países difieren a los presentes en el resto de los países, en los que los eventos tienen un grado mayor de dispersión.

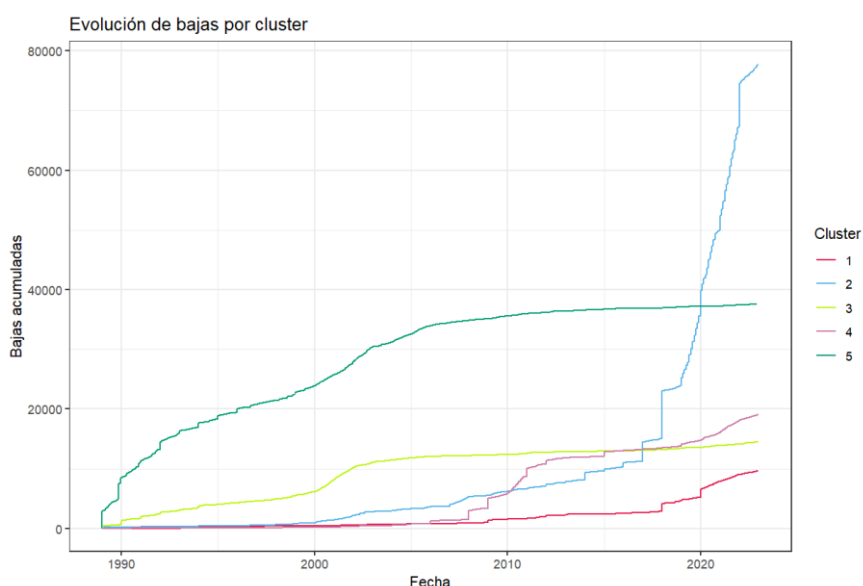
Tras el análisis visual se puede analizar el porcentaje de casos por país en cada clúster con el fin de identificar los países presentes en cada clúster. Para ello, utilizando la segmentación realizada se calcula el porcentaje de casos de cada país en cada clúster (Tabla 2).

Tabla 2. Proporción de conflictos por clúster y país.

Pais	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Venezuela	0.07%	0.42%	98.95%	0.00%	0.56%
Trinidad and Tobago	0.00%	0.00%	0.00%	0.00%	100.00%
Peru	0.00%	0.34%	25.14%	0.79%	73.73%
Paraguay	0.00%	100.00%	0.00%	0.00%	0.00%
Panama	0.00%	0.00%	0.00%	0.00%	100.00%
Nicaragua	0.00%	0.00%	0.00%	3.70%	96.30%
Mexico	13.94%	54.42%	0.60%	31.04%	0.00%
Jamaica	0.00%	20.69%	0.00%	79.31%	0.00%
Honduras	0.00%	29.73%	18.92%	51.35%	0.00%
Haiti	0.00%	20.13%	66.44%	6.04%	7.38%
Guyana	0.00%	0.00%	100.00%	0.00%	0.00%
Guatemala	0.00%	1.47%	40.46%	2.10%	55.97%
El Salvador	0.00%	0.00%	10.85%	5.43%	83.72%
Ecuador	0.00%	71.43%	3.57%	0.00%	25.00%
Dominican Republic	0.00%	0.00%	100.00%	0.00%	0.00%
Colombia	0.20%	3.11%	36.72%	0.13%	59.84%
Brazil	22.73%	58.60%	5.41%	12.63%	0.63%
Bolivia	0.00%	100.00%	0.00%	0.00%	0.00%
Argentina	100.00%	0.00%	0.00%	0.00%	0.00%

Adicionalmente, se realiza un análisis temporal de cada uno de los clústeres con el fin de entender los conflictos que principalmente han alimentado los distintos clústeres. Este permite extraer las características comunes desde un punto de vista histórico con el fin de, tras haber realizado el análisis técnico, evaluar las implicaciones de los resultados para el análisis de conflictos y la aportación a la prevención y disuasión de éstos.

Figura. 6. Evolución temporal de bajas por clúster.



Al ser el eje “y” un acumulativo de las bajas, hemos de fijarnos en aquellos periodos en los que la pendiente de las líneas es mayor, al representar un mayor incremento de bajas en el clúster. Con el fin de entender en mayor profundidad los clústeres, se analizarán algunos ejemplos representativos y se explicará la correlación entre las decisiones políticas y militares tomadas y la evolución de las bajas.

Como ejemplo más llamativo se observa el clúster 2, en el que a partir de 2018 se observa una escalada casi exponencial de bajas. El clúster recoge aquellos eventos no estatales que se dan en localizaciones no urbanas entre grupos con organización formal. No sorprende que dicha escalada muestre el recrudecimiento en la guerra de los cárteles mexicanos motivado, entre otras razones, por:

- El cambio en la política de seguridad en México (Hernández, G., 2020) con la llegada del presidente López Obrador al gobierno.
- La fragmentación de los principales cárteles, generando luchas por los territorios estratégicos y las rutas de la droga hacia los EE.UU. (Ruan, P. U., 2021)
- El incremento en la competencia por el mercado de las drogas en los EE.UU, debido principalmente a la introducción del fentanilo . (Gómez, M. L. P., 2020)

De este análisis se puede inferir que efectivamente son eventos entre grupos con organización formal (cárteles como el del Jalisco Nueva Generación o el de Sinaloa (Redaelli, C., 2021) o que la lucha por el territorio y las rutas implica la deslocalización de la violencia de las grandes ciudades (Alvear et al., 2022).

Junto con el clúster 2 se analiza la evolución del clúster 3. Este resulta de especial interés puesto que recoge específicamente eventos de violencia unilateral. En la línea de tendencia se observa un incremento muy leve hasta el año 2000 y, en comparación con el resto de valores de esa serie, un crecimiento fuerte hasta el año 2005, cuando se vuelve a estabilizar. El principal actor implicado en la violencia contra civiles son las Autodefensas Unidas de Colombia (AUC).

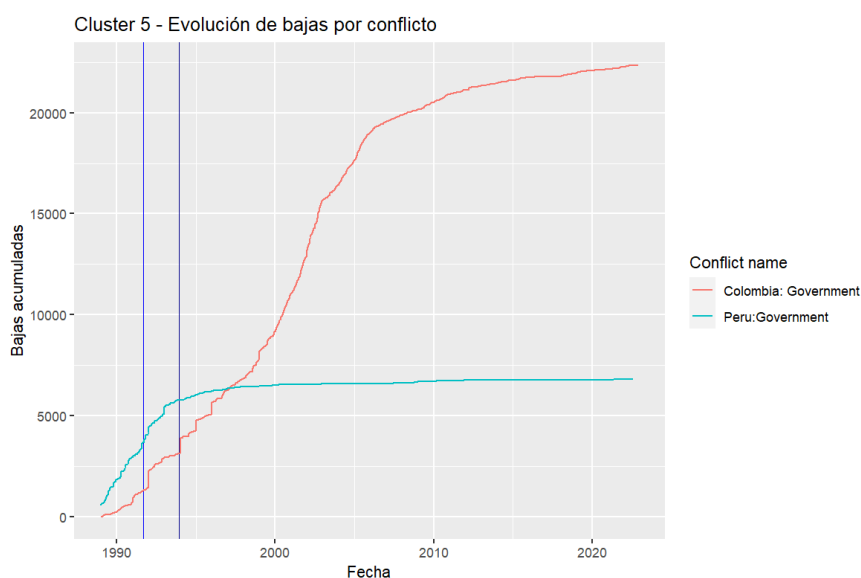
Dicha organización incrementa su nivel de violencia contra los civiles a partir del año 2000 debido a su confrontación con las FARC y el ELN, que lleva a ejercer la represión contra colaboradores, comunidades o simpatizantes de estos grupos. A eso se ha de añadir que sus relaciones con las fuerzas de seguridad les brindaron, en esos años, de una mayor capacidad e impunidad para perpetrar la violencia (Quevedo, N., 2015).

En cuanto al fin de la escalada de violencia y tras negociaciones con el gobierno de Álvaro Uribe se estableció un marco legal llamado Ley de Justicia y Paz (Cortés A., 2003). Dicho marco legal incluía incentivos legales y beneficios para los miembros que se desmovilizaran. De esta manera, las AUC comenzaron su desmovilización en el año 2003, terminando en el 2006. De esta manera, se puede observar a través del

análisis del gráfico el efecto directo de dicha desmovilización en la tendencia del clúster 3.

Por último, en el clúster 5 se perciben dos periodos con tendencia creciente. El primero correspondiente al periodo entre 1990 y 1995 y el segundo entre el 1997 y el 2003. Teniendo en cuenta que dicho clúster recoge conflictos en los que uno de los lados es un actor estatal, se observa que la subida inicial se debe mayormente al conflicto entre el Gobierno de Perú y la organización terrorista Sendero Luminoso (Rénique, J. L., 2003) y, la posterior, a enfrentamientos entre el gobierno colombiano y otros grupos armados.

Figura. 7. Evolución temporal principales conflictos (Clúster 5).



De estos dos conflictos se analizará únicamente el caso peruano debido a que, a pesar de que el inicio de este se sitúa en fechas anteriores a la serie de eventos recogida por el UCDP, la finalización se observa claramente en el gráfico a partir del año 1994. Esto permite realizar un análisis más certero de las causas que provocaron su final.

Para este caso se identifican dos hechos fundamentales que pudieron marcar el final de la insurgencia, representados en el gráfico con sendas barras verticales:

- La promulgación del Decreto Legislativo N°741 por parte del presidente Alberto Fujimori en 1991, cuya finalidad fue dar estatus legal a los grupos de autodefensa (denominados Comités de Autodefensa).
- La captura de Abimael Guzmán, principal líder de Sendero Luminoso, y la posterior firma del acuerdo de paz con el gobierno en diciembre de 1993.

Al estar estos hechos situados en años distintos, nos permiten comprobar el efecto de ambas aproximaciones a la resolución del conflicto. Como se puede comprobar en el gráfico, la aproximación que tuvo un mayor impacto en la finalización del conflicto fue la segunda, consistente en la captura del líder de la organización terrorista y la posterior firma del acuerdo de paz (Pinedo, M. E., 2021).

A través de estos ejemplos se ha podido comprobar la utilidad del análisis temporal de los datos de los clústeres, al permitir identificar los patrones históricos y factores que llevan a la finalización de los conflictos.

IV. CONCLUSIONES

En las secciones anteriores de este documento, se ha observado que la clasificación de los eventos violentos pertenecientes a los diversos conflictos latinoamericanos se ha mostrado como una parte importante de cara a su comprensión y resolución. Debido a la naturaleza volátil de los conflictos y al hecho de que los grupos involucrados en ellos se adaptan a las respuestas de los gobiernos y a las situaciones económicas y geopolíticas, la identificación de patrones comunes presentes en los eventos se convierte en una prioridad de cara a la prevención y resolución de los conflictos. Es, por tanto, fundamental comprender su naturaleza para ser capaces de aplicar las herramientas adecuadas.

La naturaleza de los eventos violentos en América Latina desde el año 1989 ha ido cambiando en los distintos países que la conforman, a la par que se han ido incrementando los conflictos y las bajas que provocan. Por ello, el análisis de cada evento a partir de los datos proporcionados por las fuentes recopiladas es prioritario para el estudio de los conflictos, puesto que permite observar una serie de

características comunes entre eventos que pueden ser lejanos, pero entre los que existen correlaciones intrínsecas identificadas por el clustering.

La detección de las mencionadas correlaciones y su consecuente clasificación en clústeres entre los eventos violentos permite aplicar a dichos eventos similares aproximaciones orientadas a su finalización que han demostrado ser más efectivas cuando se han utilizado en otros casos previos. En resumen, las estrategias utilizadas que integran el análisis de clústeres expuesto en este trabajo son más eficaces que las aproximaciones tradicionales al realizar un análisis más pormenorizado de los eventos. también se ha observado que el uso de la técnica de clusterización permite descartar variables que no aportan información de cara a la clasificación de los eventos.

Es por todo ello por lo que, teóricamente, se podría estructurar una respuesta a un conflicto utilizando el algoritmo para guiar en su resolución. De esta manera, se podría clasificar los eventos presentes en un conflicto, permitiendo dicho análisis una mejor comprensión de este y de su evolución en el tiempo. En este artículo se ha utilizado la clusterización para agrupar los eventos violentos que forman parte de distintos conflictos en Latinoamérica desde 1989 hasta la actualidad, basándonos en diversas variables que reflejan tanto el evento como la situación social y económica en el que se desarrolla, de cara a clasificar dichos eventos para su mejor comprensión. La fortaleza de los datos utilizados radica en la minuciosidad de la recogida, lo que garantiza su calidad y su pervivencia en el tiempo, facilitando de esta manera el seguimiento de conflictos en curso.

En cuanto a los clústeres detectados se ha observado que la única categoría que permite diferenciarlos de forma clara es el tipo de conflicto con la descripción de la organización de los grupos implicados, siendo esta la variable clave. Sin embargo, los clústeres 1, 2 y 4, que comparten tipología, se pueden diferenciar claramente a través del resto de variables, especialmente la referente al tipo de localización de los eventos y al Índice de Desarrollo Humano de las zonas en las que suceden. De esta forma, y observando la representación geoespacial, se observa la presencia de los

eventos del cluster 1 en los grandes núcleos poblacionales y la presencia casi constante de los eventos del cluster 2 en las regiones del este de Brasil y a lo largo de México.

Para una mejor comprensión del efecto de dicha variable en los clusters se ha de analizar el papel del estado como actor de referencia para el análisis de los conflictos armados y su resolución (Wallensteen, P., 2018). El estado juega un papel diferencial al ser, de acuerdo con la mayoría de las definiciones de las ciencias políticas, el único actor legitimado para el uso de la violencia en una sociedad.

Dicha consideración, sin embargo, no es monolítica para los diversos estados latinoamericanos, como ha podido comprobarse en el análisis.

En Latinoamérica pueden observarse situaciones en las que la dinámica establecida cambia, puesto que el estado es demasiado débil para imponer el monopolio de la fuerza. En esas ocasiones, se encuentran regiones de los estados en las que organizaciones no estatales se comportan como un cuasi estado, tratando de mantener el monopolio de la fuerza.

Dos ejemplos pueden servir para ilustrar la problemática que se comenta. El análisis detallado de dichos casos se ha llevado a cabo en los apartados dedicados al análisis geoespacial y temporal de los eventos, donde se observaba que la densidad de eventos era especialmente elevada fundamentalmente en dos países, Colombia y México.

El primer ejemplo es Colombia, donde la mayor parte de los enfrentamientos entre el estado y actores no estatales implican al grupo Marxista colombiano FARC (Fuerzas Armadas Revolucionarias de Colombia). Inicialmente, y con un carácter más marcado de insurgencia, se desarrolló y financió protegiendo el negocio de la droga en las regiones que controlaba. Esta etapa inicial queda reflejada en el clúster 5, donde los enfrentamientos del grupo son contra el gobierno, con un nivel de virulencia mayor.

Posteriormente, tras alejarse de sus orígenes ideológicos y convertirse en una “insurgencia comercial” (Metz, 1993) fue desafiado por otros grupos y adquirió, en los territorios que dominaba, ese papel cuasi estatal que se comentaba anteriormente. Es por ello que los eventos de esa segunda etapa pertenecen a los clústeres 2 y 3, con un peso más notable en el clúster 3, que recoge casos de violencia unilateral contra civiles.

Otro ejemplo que queda reflejado en los datos es el caso de México. Además de por su presencia en la mayoría de los clústeres debido al volumen de eventos, la evolución temporal de cada clúster nos permite observar la evolución del conflicto entre el estado mexicano y los distintos cárteles.

Tras un periodo temporal en el que los conflictos en México incluyen al gobierno mexicano, el incremento de la competición entre cárteles a partir del 2015 (Dunn & Balzacq, 2016) queda reflejado en el clúster 2 con la escalada de bajas que ha sido analizada de forma pormenorizada anteriormente.

Para concluir, podemos afirmar que el análisis aplicado a los conflictos en América Latina nos brinda una visión más clara de las similitudes y diferencias entre los ellos y entre los actores presentes desde el año 1989. También nos permite analizar la evolución de estos y su categorización en base al papel que juegan los actores en dicho conflicto. Esta información puede ser valiosa para el desarrollo de estrategias de seguridad más efectivas y para mejorar la prevención de conflictos violentos en la región, priorizando los recursos y esfuerzos en aquellos lugares y situaciones donde son más necesarios

V. BIBLIOGRAFÍA

- [1] Alvear, M. G., Giraldo, A., Valdez, S. & Ramírez, C. (2022). Disputa por el territorio, crimen organizado y autocorrelación espacial de homicidios en Veracruz, México,

- 2001-2018. *Revista Colombiana de Sociología*, 45(2), pp 223-251. <https://doi.org/10.15446/rcs.v45n2/95919>
- [2] Blair, R. A., & Sambanis, N. (2020). Forecasting civil wars: Theory and structure in an age of “Big Data” and machine learning. *Journal of Conflict Resolution*, 64(10), 1885-1915. <https://doi.org/10.1177/0022002720918923>
- [3] Bowsby, D., Chenoweth, E., Hendrix, C., & Moyer, J. D. (2020). The future is a moving target: predicting political instability. *British Journal of Political Science*, 50(4), 1405-1417. <https://doi.org/10.1017/S0007123418000443>
- [4] Buhaug, H., & von Uexkull, N. (2021). Vicious circles: violence, vulnerability, and climate change. *Annual Review of Environment and Resources*, 46, 545-568. <https://doi.org/10.1146/annurev-environ-012220-014708>
- [5] Chen, C., Roberts, J., Adhikari, S., Asal, V., Beardsley, K., Gonzalez, E., ... & van Wijk, A. (2022). Tipping Points: Challenges in Analyzing International Crisis Escalation. *International Studies Review*, 24(3), viac024. <https://doi.org/10.1093/isr/viac024>
- [6] Cortés, A., et al. "Desarme, desmovilización y reintegración, DDR: una introducción para Colombia." *Cuaderno de análisis* 1.13 (2013): 1-88.
- [7] Dunn Cavelty, M., & Balzacq, T. (Eds.). (2016). *Routledge Handbook of Security Studies* (2nd ed.). Routledge. <https://doi.org/10.4324/9781315753393>
- [8] Duursma, A. (2022). Peacekeeping, Mediation, and the Conclusion of Local Ceasefires in Non-State Conflicts. *Journal of Conflict Resolution*, 00220027221148132. <https://doi.org/10.1177/00220027221148132>
- [9] Ge, Q., Hao, M., Ding, F., Jiang, D., Scheffran, J., Helman, D., & Ide, T. (2022). Modelling armed conflict risk under climate change with machine learning and time-series data. *Nature communications*, 13(1), 2839. <https://doi.org/10.1038/s41467-022-30356-x>
- [10] Gleditsch, K. S., & Ward, M. D. (2013). Forecasting is difficult, especially about the future: Using contentious issues to forecast interstate disputes. *Journal of Peace Research*, 50(1), 17-31. <https://doi.org/10.1177/0022343312449033>
- [11] Goldstone, J. A., Bates, R. H., Epstein, D. L., Gurr, T. R., Lustik, M. B., Marshall, M. G., Ulfelder, J. & Woodward, M. (2010). A global model for forecasting political instability. *American Journal of Political Science*, 54(1), 190-208. <https://doi.org/10.1111/j.1540-5907.2009.00426.x>

- [12] Hegre, H., Karlsen, J., Nygård, H. M., Strand, H., & Urdal, H. (2013). Predicting armed conflict, 2010–2050. *International Studies Quarterly*, 57(2), 250-270. <https://doi.org/10.1111/isqu.12007>
- [13] Hegre, H., Nygård, H. M., & Landsverk, P. (2021). Can we predict armed conflict? How the first 9 years of published forecasts stand up to reality. *International Studies Quarterly*, 65(3), 660-668. <https://doi.org/10.1093/isq/sqaa094>
- [14] Hernández, G. (2020). La política de seguridad en México entre 2012 y 2018. De la confrontación a la prevención del delito. *Espiral (Guadalajara)*, 27(77), 43-79. <https://doi.org/10.32870/eees.v27i77.7082>
- [15] Hyde, S. D., & Saunders, E. N. (2020). Recapturing regime type in international relations: Leaders, institutions, and agency space. *International Organization*, 74(2), 363-395. <https://doi.org/10.1017/S0020818319000365>
- [16] Joshi, M. (2020). A comparative analysis of one-sided violence and civil war peace agreement implementation. *Stability: International Journal of Security and Development*, 9(1). <https://doi.org/10.5334/sta.774>
- [17] Khokhlov, N., & Korotayev, A. (2022). Internet, political regime and terrorism: A quantitative analysis. *Cross-Cultural Research*, 56(4), 385-418. <https://doi.org/10.1177/10693971221085343>
- [18] Lin, K., Yang, J., Giang, N., & Joye, C. (2021). Techniques for Forecasting Global Conflicts. <http://dx.doi.org/10.2139/ssrn.3939995>
- [19] Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2022). cluster: Cluster Analysis Basics and Extensions. R package version 2.1.4 — For new features, see the 'Changelog' file (in the package source), <https://CRAN.R-project.org/package=cluster>
- [20] Metz, S. (1993) *The Future of Insurgency*, Carlisle, PA: Strategic Studies Institute U.S. Army War College.
- [21] Pastor, M. L. (2020). El coronavirus golpea las finanzas de los cárteles mexicanos. Documento Informativo IEEE 15//2020. http://www.ieee.es/Galerias/fichero/docs_informativos/2020/DIEEEI15_2020LUIPAS_COVIDMexico.pdf
- [22] Pettersson, T., Davies, S., Deniz, A., Engström, G., Hawach, N., Höglbladh, S., & Öberg, M. S. M. (2021). Organized violence 1989–2020, with a special emphasis on Syria.

- Journal of Peace Research, 58(4), 809-825.
<https://doi.org/10.1177/00223433211026126>
- [23] Pinedo, M. E. (2021). La política después de la “guerra”. Los acuerdistas de Sendero Luminoso (1993-2020). *Eirene Estudios de Paz y Conflictos*, 4(7).
<https://orcid.org/0000-0003-3137-2424>
- [24] Quevedo, N. (9 de mayo de 2015). «Los hilos sueltos de “Macaco”». *El Espectador* (Colombia). Consultado el 20 de mayo de 2023.
<https://www.elespectador.com/investigacion/los-hilos-sueltos-de-macaco-article-559630/>
- [25] Quinn, J., Mason, T. D., Kirisci, M., & Sharif, S. (2023). Proto-Insurgency, Repression-Driven Contagion, and Civil War Onset. *Defence and Peace Economics*, 1-21.
<https://doi.org/10.1080/10242694.2023.2183346>
- [26] R Core Team, “R: A language and environment for statistical computing”. R Foundation for Statistical Computing, Vienna, Austria, 2021, <https://www.R-project.org/>
- [27] Redaelli, C. (2021). The War against Drugs: Challenges for International Humanitarian Law. *Anuario Iberoamericano de Derecho Internacional Humanitario*.
<http://dx.doi.org/10.2139/ssrn.3958287>
- [28] Rénique, J. L. (2003). La voluntad encarcelada: las luminosas trincheras de combate de Sendero Luminoso del Perú. IEP. Lima.
- [29] Rokach, L., & Maimon, O. (2005). “Clustering methods” in *Data mining and knowledge discovery handbook* (pp. 321-352). Springer, Boston, MA.
- [30] RStudio Team. “RStudio: Integrated Development for R”. RStudio, PBC, Boston, MA URL, 2020, <http://www.rstudio.com/>
- [31] Ruan, P. U. (2021). El crimen en América Latina: desorden, fragmentación y transnacionalidad. *Análisis del Real Instituto Elcano (ARI)*, (103), 1.
- [32] Russett, B. M., & Oneal, J. R. (2001). *Triangulating peace: Democracy, interdependence, and international organizations*. WW Norton & Company Incorporated.
- [33] Sundberg, R., & Melander, E. (2013). Introducing the UCDP georeferenced event dataset. *Journal of Peace Research*, 50(4), 523-532.
<https://doi.org/10.1177/0022343313484347>

- [34] Wallensteen, P. (2018) *Understanding Conflict Resolution*. 5th edn. SAGE Publications. Available at: <https://www.perlego.com/book/1432011/understanding-conflict-resolution-pdf> (Accessed: 1 May 2023).
- [35] Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- [36] Williford, G. W., & Atkinson, D. B. (2020). A Bayesian forecasting model of international conflict. *The Journal of Defense Modeling and Simulation*, 17(3), 235-242. <https://doi.org/10.1177/1548512919827659>