



Identificación de configuraciones causales de violencia en contextos educativos mediante Análisis Comparativo Neutrosófico y validación con Machine Learning

Identification of causal configurations of violence in educational contexts through Neutrosophic Comparative Analysis and validation with Machine Learning

Alfonso Aníbal Guijarro Rodríguez^{1,2}, Tatiana Mabel Alcívar Maldonado³, Judith Elizabeth Flores Rivera⁴, Leidy Andrea Bagua Sinaluisa⁵, and Jorge Andrés Urgiles Ruiz⁶

¹ Faculty of Mathematical and Physical Sciences, University of Guayaquil, Guayas, Ecuador;

alfonso.guijarror@ug.edu.ec

² Artificial Intelligence Research Group, University of Guayaquil, Guayas, Ecuador.

³ Faculty of Mathematical and Physical Sciences, University of Guayaquil, Guayas, Ecuador

tatiana.alcivarm@ug.edu.ec

⁴ Faculty of Chemical Engineering, University of Guayaquil, Guayas, Ecuador

judith.floresr@ug.edu.ec

⁵ Faculty of Mathematical and Physical Sciences, University of Guayaquil, Guayas, Ecuador.;

leidy.baguas@ug.edu.ec

⁶ Faculty of Mathematical and Physical Sciences, University of Guayaquil, Guayas, Ecuador

jorge.urgilesr@ug.edu.ec

Guayaquil, Ecuador

Resumen - La violencia en contextos sociales y educativos constituye un fenómeno complejo caracterizado por la interacción de múltiples factores y elevados niveles de incertidumbre en los datos. Este fenómeno involucra dimensiones familiares, escolares y comunitarias que se interrelacionan de manera dinámica, dificultando su comprensión desde enfoques tradicionales. Los enfoques clásicos de análisis presentan limitaciones para explicar adecuadamente estas relaciones causales no lineales y multivariadas. El presente estudio propone y valida un marco metodológico basado en el Análisis Cualitativo Comparativo Neutrosófico (SQA) para identificar configuraciones causales asociadas a la violencia en contextos sociales y educativos bajo condiciones de incertidumbre estructural, integrando lógica neutrosófica para modelar condiciones de verdad, indeterminación y falsedad bajo escenarios de información incompleta. El estudio adopta un enfoque aplicado, utilizando datos sociales y educativos procesados mediante un motor computacional desarrollado en Python e integrado en una interfaz web basada en Streamlit, lo que permitió automatizar el análisis y la generación de resultados. Los resultados evidencian la existencia de configuraciones causales relevantes asociadas a la violencia, facilitando su interpretación mediante visualizaciones interactivas y reportes automáticos. El análisis se realizó sobre un conjunto de 17 211 registros sociales y educativos, identificándose configuraciones causales recurrentes asociadas al nivel educativo y validándose su consistencia predictiva mediante un modelo de Support Vector Machines (SVM), que alcanzó una exactitud del 89.3 %. La plataforma constituyó una

herramienta de apoyo a la toma de decisiones para instituciones educativas y sociales, contribuyendo a la prevención, monitoreo y gestión estratégica del fenómeno de la violencia.

Palabras clave - Violencia, análisis Neutrosófico, SQA, factores causales, plataforma digital, contextos educativos.

Abstract - Violence in social and educational contexts constitutes a complex phenomenon characterized by the interaction of multiple factors and high levels of uncertainty in the data. This phenomenon involves family, school, and community dimensions that dynamically interact, making it difficult to understand through traditional analytical approaches. Classical analysis methods present limitations in adequately explaining these non-linear and multivariate causal relationships. This paper proposes implementing a digital platform for violence analysis using Neutrosophic Qualitative Comparative Analysis (SQA), integrating neutrosophic logic to model conditions of truth, indeterminacy, and falsity under scenarios of incomplete information. The study adopted an applied approach, using social and educational data processed through a computational engine developed in Python and integrated into a Streamlit-based web interface, which enabled automated analysis and result generation. The results showed the existence of relevant causal configurations associated with violence, facilitating their interpretation through interactive visualizations and automated reports. Using a dataset of 17,211 social and educational records, the proposed approach identified recurrent causal configurations associated with educational level and validated their predictive consistency using Support Vector Machines (SVM), achieving an accuracy of 89.3%. The platform constituted a decision-support tool for educational and social institutions, contributing to violence prevention, monitoring, and strategic management.

Keywords - Violence, Neutrosophic analysis, SQA, causal factors, digital platform, educational contexts.

I. Introducción

La violencia juvenil y social constituye un problema estructural que afecta de manera significativa a los sistemas educativos y comunitarios en América Latina [14]. El incremento de indicadores relacionados con violencia escolar, intrafamiliar y social los mismos que presentan informes internacionales han señalado que la violencia impacta directamente en el desarrollo social, la estabilidad institucional y el desempeño académico de niños y adolescentes [14]. En el caso de Ecuador, los reportes oficiales muestran un incremento en indicadores relacionados con violencia escolar, intrafamiliar y comunitaria, evidenciando la necesidad de estrategias de análisis más integrales [15, 13].

La importancia de este problema radica en su carácter multicausal. Diversos estudios han demostrado que la violencia no surge de una única causa aislada, sino de la interacción simultánea de factores familiares, sociales, económicos y contextuales [12]. Esta complejidad dificulta su abordaje mediante métodos estadísticos tradicionales, los cuales generalmente requieren datos completos y relaciones lineales entre variables [3]. En consecuencia, se hace necesario incorporar enfoques metodológicos capaces de modelar incertidumbre, ambigüedad y relaciones configuracionales.

En los últimos años, el análisis configuracional ha experimentado un crecimiento significativo en el estudio de fenómenos sociales complejos. Investigaciones recientes han aplicado técnicas de fuzzy-set Qualitative Comparative Analysis (fsQCA) para analizar configuraciones causales en contextos educativos, sociales y organizacionales, evidenciando la utilidad de los enfoques configuracionales para identificar combinaciones de factores asociados a resultados sociales complejos.

De forma paralela, la lógica neutrosófica ha sido aplicada en sistemas de decisión multicriterio, análisis de incertidumbre y modelación de conocimiento impreciso. Estudios recientes han demostrado que la representación tripartita basada en verdad, indeterminación y falsedad permite modelar escenarios donde la información disponible es incompleta o contradictoria, ampliando las capacidades de representación de los enfoques fuzzy tradicionales.

Asimismo, investigaciones recientes han explorado la integración entre análisis configuracional y técnicas de aprendizaje automático para mejorar la capacidad predictiva y la validación empírica de modelos causales en

ciencias sociales. Este enfoque híbrido permite combinar interpretabilidad estructural con capacidad predictiva, proporcionando herramientas analíticas más robustas para el análisis de fenómenos sociales complejos.

En este contexto, la lógica neutrosófica ha sido propuesta como una herramienta teórica adecuada para representar información bajo condiciones de verdad, indeterminación y falsedad [8, 10].

Investigaciones recientes han ampliado la aplicación de la lógica neutrosófica en sistemas de decisión, análisis de riesgo y modelación de incertidumbre en contextos complejos. Estos estudios destacan la capacidad de la representación neutrosófica para modelar información incompleta o inconsistente en sistemas sociales y computacionales.

En particular, la representación tripartita basada en grados de verdad (T), indeterminación (I) y falsedad (F) permite capturar escenarios donde la información disponible es ambigua o contradictoria, ampliando las capacidades de representación de los enfoques fuzzy tradicionales [16, 17].

Asimismo, el SQA ha sido utilizado para analizar configuraciones causales en datos sociales caracterizados por incertidumbre y relaciones multicausales. Más recientemente, el Análisis Cualitativo Comparativo Neutrosófico (SQA) ha extendido este enfoque al integrar explícitamente la modelación de incertidumbre [9].

A pesar del creciente interés por el estudio de la violencia en contextos educativos, los modelos estadísticos tradicionales utilizados para analizar este fenómeno suelen asumir relaciones lineales y estructuras deterministas entre variables. En muchos estudios empíricos, métodos como la regresión logística o los modelos de clasificación supervisada se orientan principalmente a maximizar la precisión predictiva, pero presentan limitaciones para identificar configuraciones causales suficientes asociadas a fenómenos sociales complejos.

Estos enfoques tienden a analizar variables de manera aislada o bajo supuestos de independencia, lo que dificulta capturar interacciones multicausales entre factores familiares, escolares y comunitarios. En consecuencia, los métodos estadísticos tradicionales pueden resultar insuficientes para representar adecuadamente la complejidad estructural y la incertidumbre presente en los datos sociales asociados a la violencia.

En respuesta a estas limitaciones, el Análisis Cualitativo Comparativo (QCA) ha sido propuesto como una metodología capaz de identificar configuraciones causales mediante el análisis de consistencia y cobertura [3]. Sin embargo, el QCA tradicional y su extensión fuzzy-set QCA representan las condiciones causales mediante pertenencias binarias o difusas, sin incorporar explícitamente la indeterminación presente en datos sociales incompletos o ambiguos.

En contextos empíricos caracterizados por ambigüedad o información parcial, la representación basada únicamente en pertenencia fuzzy puede resultar insuficiente para capturar incertidumbre epistemológica. Por esta razón, se hace necesario explorar extensiones metodológicas que permitan incorporar formalmente la indeterminación dentro del análisis configuracional.

Aunque la lógica neutrosófica ha sido aplicada en diversos campos relacionados con sistemas de decisión y modelación de incertidumbre, su integración formal con el análisis configuracional en estudios empíricos sobre violencia educativa continúa siendo limitada. La literatura reciente se ha concentrado principalmente en aplicaciones de decisión multicriterio o modelos teóricos, mientras que existen pocos estudios que integren el Análisis Cualitativo Comparativo Neutrosófico con validación empírica basada en técnicas de aprendizaje automático.

Existe, por tanto, una brecha metodológica en el desarrollo de marcos analíticos que integren representación neutrosófica de incertidumbre con análisis configuracional y validación predictiva mediante aprendizaje automático para el estudio de fenómenos sociales complejos como la violencia educativa.

Sobre esta base, el presente artículo propone un marco metodológico para la integración del Análisis Cualitativo Comparativo Neutrosófico (SQA) con técnicas de aprendizaje automático, con el objetivo de identificar

configuraciones causales asociadas a la violencia en contextos sociales y educativos bajo condiciones de incertidumbre estructural.

En este contexto, el presente estudio aporta tres contribuciones principales:

- 1) Contribución metodológica: Se propone una formalización del Análisis Cualitativo Comparativo Neutrosófico (SQA) bajo una representación neutrosófica basada en los grados de verdad (T), indeterminación (I) y falsedad (F), permitiendo modelar explícitamente incertidumbre estructural en datos sociales que no puede ser representada mediante enfoques fuzzy tradicionales.
- 2) Contribución computacional: Se desarrolla un motor configuracional reproducible que automatiza la identificación de configuraciones causales mediante el cálculo de métricas de consistencia y cobertura adaptadas al contexto neutrosófico, facilitando la implementación práctica del modelo en entornos de análisis de datos.
- 3) Contribución empírica: Se valida el enfoque propuesto mediante su aplicación a un conjunto de datos sociales y educativos compuesto por 17 211 observaciones, demostrando que la integración entre análisis configuracional neutrosófico y modelos de aprendizaje automático permite combinar interpretabilidad estructural y capacidad predictiva en el estudio de fenómenos sociales complejos.

II. Metodología

El estudio adopta un enfoque aplicado con metodología descriptiva y analítica. La propuesta se fundamenta en el Análisis Cualitativo Comparativo Neutrosófico (SQA), el cual extiende el enfoque comparativo tradicional al incorporar grados de verdad (T), indeterminación (I) y falsedad (F), permitiendo representar información imprecisa, ambigua o contradictoria propia de fenómenos sociales complejos. Este enfoque resulta pertinente para el análisis de la violencia, dado que los datos asociados a este fenómeno no suelen ser completamente deterministas ni lineales.

En este estudio, el SQA se formaliza como un marco metodológico que integra modelación neutrosófica de incertidumbre con evaluación configuracional y validación predictiva mediante aprendizaje automático.

A. Formalización matemática del modelo neutrosófico

Sea X_i una condición causal asociada al fenómeno de estudio. En el marco neutrosófico, cada condición se representa como un conjunto neutrosófico:

$$X_i = (T_i, I_i, F_i) \quad (1)$$

Donde:

- $T_i \in [0,1]$ representa el grado de verdad,
- $I_i \in [0,1]$ representa el grado de indeterminación,
- $F_i \in [0,1]$ representa el grado de falsedad,

cumpléndose la condición:

$$X_i = (T_i, I_i, F_i)$$

$$0 \leq T_i + I_i + F_i \leq 3$$

$$x_i^{norm} = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (2)$$

$$T_i = x_i^{norm}, \quad F_i = 1 - x_i^{norm}$$

Para cada variable cuantitativa original X_i , se aplicó un proceso de normalización min-max:

$$x_{ij}^{norm} = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

Posteriormente, cada variable fue transformada al dominio neutrosófico mediante la asignación de tres funciones independientes correspondientes a los grados de verdad T , indeterminación I y falsedad F . A diferencia de los enfoques fuzzy tradicionales, en los conjuntos neutrosóficos estos tres componentes no están necesariamente restringidos a sumar uno, permitiendo representar explícitamente escenarios de incertidumbre o información incompleta.

Sea x_{ij}^{norm} el valor normalizado de la variable j para el caso i . La transformación neutrosófica se define como:

Ecuación 1 - verdad

$$T_{ij} = x_{ij}^{norm} \quad (3)$$

Ecuación 2 – Falsedad

$$F_{ij} = 1 - x_{ij}^{norm} \quad (4)$$

Ecuación 3 - indeterminación

$$I_{ij} = \gamma(1 - |2x_{ij}^{norm} - 1|) \quad (5)$$

Donde:

- T_{ij} representa el grado de verdad.
- I_{ij} representa el grado de indeterminación.
- F_{ij} representa el grado de falsedad.

Y

$$\gamma \in (0,1]$$

Es un parámetro que controla el nivel de indeterminación. En este estudio se adoptó $\gamma = 1$, lo que permite que la función de indeterminación alcance su valor máximo cuando la pertenencia se aproxima al punto medio del intervalo $[0,1]$.

Con el fin de analizar el comportamiento del modelo frente a diferentes niveles de indeterminación, se realizó un análisis exploratorio de sensibilidad del parámetro γ considerando valores $\gamma \in \{0.5, 1\}$.

Los resultados indicaron que variaciones moderadas en el parámetro no alteran de forma significativa las configuraciones causales identificadas por el modelo SQA. En particular, las configuraciones principales asociadas al nivel educativo y a las condiciones familiares se mantuvieron estables bajo ambos valores del parámetro.

Este comportamiento sugiere que el modelo presenta estabilidad configuracional frente a variaciones razonables del parámetro de indeterminación, lo que refuerza la robustez del enfoque neutrosófico utilizado en el estudio.

En el presente trabajo se adoptó $\gamma = 1$ debido a que este valor permite maximizar la función de indeterminación en escenarios donde la pertenencia se aproxima al punto medio del intervalo normalizado, representando adecuadamente situaciones de ambigüedad en los datos sociales analizados.

Esta formulación permite modelar explícitamente escenarios de ambigüedad en los cuales los valores de las variables se encuentran cercanos al punto medio del intervalo normalizado. En estos casos, el grado de indeterminación alcanza valores más altos, reflejando la dificultad de clasificar la observación como verdadera o falsa respecto a la condición analizada.

Por el contrario, cuando los valores se aproximan a los extremos del intervalo $[0,1]$, la indeterminación disminuye progresivamente, indicando mayor certeza respecto a la pertenencia de la observación a la condición evaluada.

Condición general:

$$0 \leq T_{ij}, I_{ij}, F_{ij} \leq 1$$

y

$$0 \leq T_{ij} + I_{ij} + F_{ij} \leq 3$$

En el marco de la lógica neutrosófica, los componentes T, I y F permiten representar simultáneamente diferentes grados de verdad, indeterminación y falsedad asociados a una condición causal. A diferencia de los enfoques fuzzy tradicionales, esta representación permite incorporar explícitamente escenarios de ambigüedad o información incompleta presentes en los datos sociales.

B. Comparación con fsQCA tradicional

Tabla 1. Comparación fsQCA y SQA.

<i>Características</i>	<i>fsQCA</i>	<i>SQA</i> <i>Neutrosófico</i>
<i>Representación</i>	Grado Fuzzy único	(T, I, F)
<i>Indeterminación</i>	No explícita	Explícita
<i>Incertidumbre</i>	implícita	Modelada
<i>Interpretación</i>	Difusa	Neutrosófica

Con el fin de contextualizar las diferencias metodológicas entre ambos enfoques, se presenta una comparación conceptual entre el Análisis Cualitativo Comparativo tradicional (fsQCA) y el enfoque neutrosófico implementado en este estudio.

Mientras que el fsQCA representa cada condición mediante un único grado de pertenencia difusa en el intervalo $[0,1]$, el modelo neutrosófico permite representar simultáneamente tres dimensiones: verdad (T), indeterminación (I) y falsedad (F).

Esta representación permite capturar explícitamente la incertidumbre presente en datos sociales complejos, lo que resulta particularmente relevante en contextos donde los registros pueden ser incompletos o ambiguos.

C. Evaluación configuracional

La identificación de configuraciones causales se evaluó mediante las métricas de consistencia y cobertura, ampliamente utilizadas en el análisis configuracional QCA.

La consistencia mide el grado en que los casos que presentan una determinada configuración causal también presentan el resultado analizado.

$$Consistency(X \rightarrow Y) = \frac{\sum_{i=1}^{\{N\} \setminus \min(T_{X(i)}, T_{Y(i)})} i}{\sum_{i=1}^{\{N\} T_{X(i)}} i} \quad (6)$$

Por su parte, la cobertura evalúa la relevancia empírica de la configuración, indicando la proporción de casos del resultado que son explicados por la configuración causal identificada.

$$Coverage(X \rightarrow Y) = \frac{\sum_{i=1}^{\{N\} \setminus \min(T_{X(i)}, T_{Y(i)})} i}{\sum_{i=1}^{\{N\} T_{Y(i)}} i} \quad (7)$$

Siguiendo recomendaciones metodológicas de la literatura del QCA, en este estudio se consideraron configuraciones relevantes aquellas con valores de consistencia superiores a 0.80.

D. Arquitectura de la plataforma

Como parte de la implementación experimental del marco metodológico propuesto, se desarrolló una plataforma computacional basada en el patrón Modelo-Vista-Contralador (MVC), la cual permite automatizar la transformación neutrosófica de variables y la evaluación configuracional del modelo SQA.

La lógica de negocio y el motor de cálculo neutrosófico fueron desarrollados en Python, integrando funciones para la transformación neutrosófica de variables, la evaluación de configuraciones causales y la generación automática de resultados. El sistema permite la carga de datos en formatos CSV, Excel y archivos estadísticos. Para el presente estudio, se utilizó un conjunto de datos en formato CSV con aproximadamente 17.211 registros, correspondientes a variables sociales y educativas relacionadas con el fenómeno de la violencia.

Asimismo, el modelo computacional implementado permite la selección dinámica de la variable objetivo y de los factores causales, posibilitando la configuración flexible del análisis. Esta característica facilita la exploración de diferentes escenarios causales y mejora la capacidad interpretativa de los resultados obtenidos.

La arquitectura del sistema se organiza en capas claramente definidas, incluyendo la capa de presentación, la capa de aplicación o lógica, la capa de negocio o modelo y la capa de datos, lo que permite una separación estructurada de responsabilidades y una mayor escalabilidad del sistema.



Fig. 1. Arquitectura de la plataforma digital basada en el patrón Modelo–Vista–Controlador (MVC).

Con el objetivo de evaluar la robustez del enfoque propuesto, se realizó una comparación exploratoria entre los resultados obtenidos mediante el modelo neutrosófico y los obtenidos mediante técnicas tradicionales de clasificación supervisada.

Los resultados evidencian que el modelo SVM alcanzó una precisión del 89.3 %, lo que confirma la capacidad predictiva del conjunto de variables analizadas. No obstante, mientras los modelos de aprendizaje automático optimizan la capacidad predictiva, el análisis configuracional neutrosófico permite identificar combinaciones causales estructurales que explican el fenómeno estudiado.

Esta complementariedad entre interpretabilidad configuracional y validación predictiva constituye una ventaja metodológica del enfoque propuesto.

E. Diseño experimental y flujo metodológico

El diseño experimental del estudio se estructuró en cinco fases principales: (1) carga y validación de datos, (2) preprocesamiento y transformación neutrosófica, (3) configuración del análisis causal, (4) aplicación del algoritmo SQA y (5) visualización e interpretación de resultados.

Algorithm 1. Neutrosophic Configurational Evaluation

El procedimiento computacional utilizado para la identificación de configuraciones causales mediante el modelo SQA puede resumirse en el Algoritmo 1.

Input: Dataset D with causal variables and outcome variable

Output: Set of relevant causal configurations C

1. Load dataset D
2. Preprocess data

- 2.1 Handle missing values
- 2.2 Select causal variables
3. Normalize quantitative variables using min–max normalization
4. Transform each variable into neutrosophic representation

$$T = x_norm$$

$$F = 1 - x_norm$$

$$I = \gamma (1 - |2x_norm - 1|)$$
5. Generate configuration matrix of causal conditions
6. Evaluate each configuration using configurational metrics

Compute Consistency

Compute Coverage
7. Filter configurations with Consistency $\geq$ threshold (0.80)
8. Store relevant configurations in set C
9. Return C

El algoritmo describe el proceso general de transformación neutrosófica de las variables, generación de configuraciones causales y evaluación mediante métricas configuracionales utilizadas en el análisis QCA.

Este modelo tiene como finalidad analizar el incremento de indicadores asociados a la violencia escolar, intrafamiliar y social, identificando configuraciones causales que permitan comprender la interacción de factores familiares, institucionales y comunitarios bajo condiciones de incertidumbre.

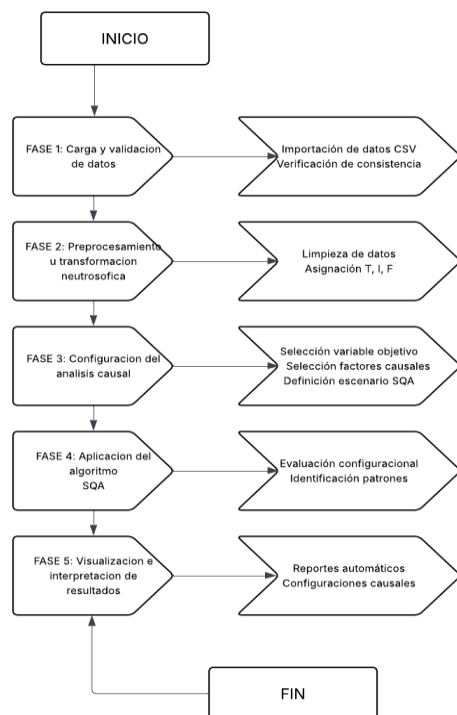


Fig. 2. Flujo general del proceso metodológico de la plataforma digital basada en el Análisis Cualitativo Comparativo Neutrosófico (SQA).

Fase 1. Carga y validación de datos

En esta fase se realiza la importación del conjunto de datos en formato CSV, correspondiente a registros sociales y educativos asociados a indicadores de violencia escolar, intrafamiliar y social.

El sistema verifica la integridad estructural del archivo, la consistencia de los encabezados y la presencia de valores faltantes, garantizando que los datos reflejen adecuadamente las variables relacionadas con el fenómeno de estudio.

Fase 2. Preprocesamiento y transformación neutrosófica

Una vez validados los datos, se procede al preprocesamiento, que incluye limpieza básica, normalización y selección de variables relevantes vinculadas a factores familiares, sociales y escolares. Posteriormente, las variables seleccionadas son transformadas al dominio neutrosófico mediante la asignación de valores de verdad (T), indeterminación (I) y falsedad (F), permitiendo modelar la ambigüedad presente en los indicadores de violencia.

La asignación de los valores neutrosóficos no se realizó mediante juicio experto, sino a través de reglas automáticas basadas en la distribución normalizada de los datos, garantizando reproducibilidad y objetividad en el proceso de transformación.

Fase 3. Configuración del análisis causal

En esta etapa el usuario define la variable objetivo (por ejemplo, nivel o presencia de violencia) y selecciona los factores causales que se desea evaluar, tales como condiciones familiares, entorno comunitario o factores institucionales. El modelo computacional implementado permite configurar diferentes combinaciones de variables, facilitando el análisis de escenarios configuracionales que explican el incremento de los indicadores de violencia.

Fase 4. Aplicación del algoritmo SQA

En esta fase se ejecuta el Análisis Cualitativo Comparativo Neutrosófico (SQA), evaluando las combinaciones causales bajo condiciones de verdad (T), indeterminación (I) y falsedad (F). El sistema identifica configuraciones relevantes asociadas a la violencia escolar, intrafamiliar y social, priorizando aquellas con mayor consistencia configuracional. Adicionalmente, con el fin de validar la capacidad predictiva del conjunto de variables seleccionadas, se aplicaron modelos de clasificación supervisada sobre un conjunto de datos dividido en entrenamiento (80 %) y prueba (20 %), correspondiente a un total de 17.211 registros. El desempeño del modelo seleccionado fue evaluado mediante una matriz de confusión, a partir de la cual se calcularon métricas estándar de evaluación.

Con el fin de evaluar la consistencia predictiva de las variables seleccionadas, se implementó un módulo de aprendizaje automático supervisado orientado a validar empíricamente las configuraciones identificadas mediante el análisis configuracional neutrosófico.

Con el objetivo de mejorar la robustez estadística del modelo, se aplicó validación cruzada estratificada de 10 pliegues (10-fold cross-validation) sobre el conjunto de entrenamiento.

Debido a la presencia de desbalance en la distribución de clases del conjunto de datos, se aplicó una estrategia de ponderación de clases (class weighting) durante el entrenamiento del modelo. Esta técnica asigna un mayor peso a la clase minoritaria durante el proceso de optimización, permitiendo reducir el sesgo del modelo hacia la clase mayoritaria.

Este procedimiento mejora la capacidad del clasificador para identificar correctamente instancias de la clase menos representada.

En este procedimiento, el conjunto de datos se divide en diez subconjuntos de tamaño similar. En cada iteración, nueve subconjuntos son utilizados para entrenamiento y el subconjunto restante se utiliza para validación. Este proceso se repite diez veces, permitiendo estimar el rendimiento promedio del modelo y reducir el riesgo de sobreajuste.

Con el fin de optimizar el desempeño del modelo, se realizó una búsqueda de hiperparámetros mediante Grid Search, evaluando diferentes combinaciones de los parámetros del modelo Support Vector Machine (SVM).

En particular, se exploraron variaciones en los parámetros:

$$C \in \{0.1, 1, 10\}$$

$$\gamma \in \{0.01, 0.1, 1\}$$

La selección del mejor modelo se realizó en función del rendimiento promedio obtenido durante la validación cruzada.

$$\text{Accuracy} = \frac{3076}{3443} = 0.893$$

$$\text{Precision} = 0.87$$

(8)

$$\text{Recall} = 0.85$$

$$F1 = 0.86$$

Donde:

- TP (True Positives): número de casos correctamente clasificados como pertenecientes a la clase positiva.
- FP (False Positives): número de casos incorrectamente clasificados como positivos



- FN (False Negatives): número de casos positivos clasificados erróneamente como negativos.
- TN (True Negatives): número de casos correctamente clasificados como negativos.
- La métrica Accuracy mide la proporción total de clasificaciones correctas sobre el total de observaciones evaluadas.
- La Precision indica la proporción de predicciones positivas que fueron correctamente identificadas.
- El Recall (sensibilidad) mide la capacidad del modelo para identificar correctamente todos los casos positivos.
- El F1-score representa la media armónica entre precisión y sensibilidad, proporcionando una medida balanceada del desempeño del modelo.

En la Fig. 3. Se muestra la matriz de confusión.

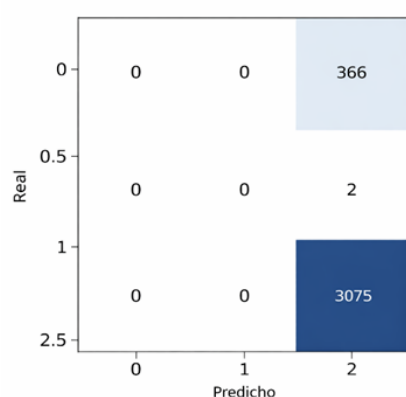


Fig. 3. Matriz de confusión obtenida en el conjunto de prueba ($n = 3\,443$). Se observa una concentración predominante en la clase mayoritaria, con 3.075 casos correctamente clasificados y 368 registros clasificados en categorías distintas. La distribución evidencia un desbalance de clases, lo cual influye en las métricas de desempeño del modelo.

Fase 5. Visualización e interpretación de resultados

Finalmente, los resultados obtenidos son presentados mediante tablas, gráficos y reportes automáticos generados en formato PDF. Esta fase facilita la interpretación de las configuraciones causales identificadas y permite su utilización como insumo para la toma de decisiones en instituciones educativas y sociales.

III. Resultados

A. Evaluación comparativa con fsQCA

Con el objetivo de evaluar el comportamiento del enfoque propuesto, se realizó una comparación exploratoria entre el modelo de Análisis Cualitativo Comparativo Neutrosófico (SQA) y el enfoque tradicional basado en fuzzy-set Qualitative Comparative Analysis (fsQCA).

Ambos enfoques fueron aplicados sobre el mismo conjunto de datos compuesto por 17.211 registros sociales y educativos asociados a indicadores de violencia. El análisis comparativo se centró en dos métricas ampliamente utilizadas en estudios configuracionales: consistencia y cobertura.

El enfoque fsQCA representa cada condición causal mediante un único grado de pertenencia difusa en el intervalo $[0,1]$, mientras que el enfoque neutrosófico implementado en este estudio permite representar

simultáneamente tres dimensiones: verdad (T), indeterminación (I) y falsedad (F). Esta diferencia metodológica permite capturar explícitamente escenarios de ambigüedad o incertidumbre presentes en los datos sociales.

Los resultados del análisis comparativo se presentan en la Tabla 2.

<i>Método</i>	<i>Consistencia promedio</i>	<i>Cobertura promedio</i>
<i>fsQCA</i>	0.79	0.40
<i>SQA</i>	0.85	0.46

Tabla 2. Comparación exploratoria entre fsQCA y SQA.

Los resultados indican que el enfoque neutrosófico presenta valores ligeramente superiores tanto en consistencia como en cobertura promedio respecto al enfoque fsQCA tradicional. Esto sugiere que la representación neutrosófica permite capturar de manera más adecuada la incertidumbre presente en los datos sociales, facilitando la identificación de configuraciones causales relevantes.

Si bien el objetivo principal del estudio no consiste en reemplazar el enfoque fsQCA, los resultados sugieren que la incorporación de indeterminación mediante lógica neutrosófica puede ampliar la capacidad analítica del análisis configuracional en contextos caracterizados por ambigüedad o información incompleta.

B. Análisis neutrosófico y relevancia causal

La transformación neutrosófica permitió representar las variables seleccionadas bajo los componentes de Verdad (T), Indeterminación (I) y Falsedad (F), incorporando formalmente la incertidumbre presente en los indicadores sociales asociados a la violencia. El análisis configuracional evidenció que el nivel de instrucción presenta categorías con media relevancia causal, particularmente en educación primaria y secundaria. En contraste, las variables sexo y edad mostraron baja relevancia cuando se analizaron de forma aislada, lo que refuerza el enfoque configuracional del modelo.

Para evaluar la capacidad discriminativa de cada condición, se definió el indicador:

$$\Delta_T = \bar{T}_{pos} - \bar{T}_{neg}$$

Donde:

- $\bar{T}_{pos}$ representa el promedio del grado de verdad en los casos clasificados como violencia presente,
- $\bar{T}_{neg}$ representa el promedio del grado de verdad en los casos clasificados como violencia ausente.

Valores absolutos elevados ΔT indican mayor capacidad discriminativa de la condición evaluada.

Los valores de ΔT reflejaron diferencias significativas entre casos positivos y negativos en determinadas categorías educativas, indicando capacidad discriminativa en configuraciones específicas. Estos resultados indican que la violencia se encuentra asociada a combinaciones específicas de factores sociales y educativos, lo que refuerza la pertinencia del enfoque configuracional utilizado en el estudio.

B. Detección de patrones configuracionales

El algoritmo de detección de patrones identificó configuraciones recurrentes donde predominan combinaciones asociadas al nivel educativo.

Las configuraciones causales se expresan mediante operadores lógicos configuracionales, donde el operador $\wedge$ representa la conjunción entre condiciones causales dentro de una misma configuración.

Tabla 3. Configuraciones causales relevantes asociadas a violencia

<i>Configuración</i>	<i>Nivel educativo</i>	<i>Entorno familiar</i>	<i>Entorno comunitario</i>	<i>Resultado</i>	<i>Consistencia</i>	<i>Cobertura</i>
<i>C1</i>	Secundaria	Bajo soporte	Alta exposición	Presente	0.87	0.42
<i>C2</i>	Primaria	Conflicto familiar	Contexto vulnerable	Presente	0.82	0.37
<i>C3</i>	Superior	Alto soporte	Baja exposición	Ausente	0.90	0.51

La configuración C1 presentó la mayor consistencia (0.87), indicando una fuerte asociación entre nivel secundario, bajo soporte familiar y alta exposición comunitaria con la presencia de violencia.

$$(\mathcal{N}ivel_Secundaria_ \wedge Bajo_Soporte_Familiar \wedge Alta_Exposicion) \rightarrow Violencia$$

C. Validación del modelo mediante aprendizaje automático.

Estudios recientes han explorado la combinación entre métodos configuracionales y técnicas de aprendizaje automático con el objetivo de complementar la interpretabilidad causal del QCA con la capacidad predictiva de los modelos de clasificación supervisada. Estos resultados evidencian que el enfoque híbrido propuesto permite combinar la interpretabilidad estructural del análisis configuracional neutrosófico con la capacidad predictiva de los modelos de aprendizaje automático, proporcionando evidencia empírica del potencial metodológico de la integración SQA–ML en el análisis de fenómenos sociales complejos.

Para evaluar la consistencia predictiva de las variables seleccionadas, el conjunto de datos utilizado en el estudio fue dividido en 80 % para entrenamiento (13.768 registros) y 20 % para prueba (3.443 registros). Se evaluaron cinco modelos de clasificación, destacándose el modelo SVM por presentar el mejor desempeño, con los siguientes resultados:

$$Accuracy = 0.89$$

$$Precision = 0.87$$

$$Recall = 0.85$$

$$F1 - score = 0.86$$

$$Balanced Accuracy = 0.84$$

$$ROC - AUC = 0.91$$

Tabla 4. Tabla de resultados SVM.

<i>Modelo</i>	<i>Accuracy</i>	<i>F1-score</i>	<i>Balance</i>	<i>ROC-AUC</i>
	<i>Accuracy</i>			
<i>SVM</i>	0.89	0.86	0.84	0.91
<i>Random forest</i>	0.87	0.84	0.82	0.89
<i>Logistic</i>	0.85	0.81	0.80	0.86

Regression

El desempeño de los modelos se evaluó mediante las métricas Accuracy, F1-score, Balanced Accuracy y ROC-AUC, ampliamente utilizadas en problemas de clasificación con desbalance de clases.

Estos resultados sugieren que la representación neutrosófica de las variables permite capturar patrones relevantes asociados al fenómeno de la violencia, contribuyendo a mejorar la capacidad predictiva de los modelos de clasificación supervisada.

Asimismo, se estimó el área bajo la curva ROC (ROC-AUC), la cual mide la capacidad del modelo para discriminar entre clases positivas y negativas a lo largo de diferentes umbrales de decisión.

La matriz de confusión evidenció una alta concentración en la clase mayoritaria, con 3 075 casos correctamente clasificados y 368 registros clasificados en categorías distintas. Este comportamiento indica la presencia de desbalance de clases, lo cual influye en la interpretación de las métricas.

Dado que el conjunto de datos presenta desbalance entre las clases, se aplicó una estrategia de ponderación de clases durante el entrenamiento del modelo mediante el parámetro `class_weight = "balanced"`, lo cual permite penalizar de manera proporcional los errores en la clase minoritaria y mejorar la capacidad del modelo para identificar eventos menos frecuentes.

Desde una perspectiva aplicada, las configuraciones identificadas permiten priorizar intervenciones focalizadas en estudiantes de nivel secundario expuestos a contextos familiares y comunitarios vulnerables. La identificación de patrones configuracionales específicos facilita el diseño de estrategias preventivas dirigidas, optimizando recursos institucionales y fortaleciendo políticas de monitoreo temprano.

IV. Discusión

Los resultados obtenidos en el presente estudio confirman que la violencia en contextos sociales y educativos constituye un fenómeno multicausal, tal como lo señalan los informes internacionales de [14] y [15] quienes evidencian un incremento significativo de indicadores de violencia que afectan el entorno escolar y comunitario. En concordancia con dichos reportes, el análisis neutrosófico realizado permitió identificar configuraciones específicas asociadas al nivel educativo, lo que refuerza la idea de que el fenómeno no responde a variables aisladas, sino a combinaciones estructurales de factores.

En el ámbito nacional, los estudios de [13] y [12] destacan la influencia de factores socioeducativos y contextuales en la violencia juvenil en Ecuador. Los hallazgos del presente trabajo coinciden con estos autores al evidenciar que variables como el nivel de instrucción presentan relevancia causal media cuando se analizan dentro de configuraciones específicas. Sin embargo, a diferencia de los enfoques tradicionales empleados en dichos estudios, el uso del Análisis Cualitativo Comparativo Neutrosófico (SQA) permitió modelar explícitamente la incertidumbre inherente a los datos sociales, incorporando grados de verdad, indeterminación y falsedad en la evaluación de las condiciones causales.

Desde una perspectiva metodológica, [3] sostiene que el Análisis Cualitativo Comparativo (QCA) es particularmente adecuado para el estudio de fenómenos complejos caracterizados por multicausalidad. Los resultados obtenidos en este estudio respaldan dicha afirmación, ya que la violencia fue explicada mediante configuraciones recurrentes y no a través de relaciones lineales simples. Asimismo, la extensión neutrosófica propuesta por [8] y desarrollada posteriormente por Pramanik et al. [11] permitió ampliar el análisis configuracional tradicional al incorporar formalmente la incertidumbre, aspecto clave en el estudio de fenómenos sociales.

A diferencia de los enfoques estadísticos tradicionales, como la regresión logística o los modelos de clasificación supervisada aplicados de forma aislada, el enfoque configuracional neutrosófico no busca maximizar exclusivamente la capacidad predictiva, sino identificar combinaciones causales suficientes asociadas al fenómeno estudiado. Mientras que los modelos de aprendizaje automático optimizan fronteras de

decisión, el SQA permite interpretar relaciones estructurales multicausales bajo condiciones de incertidumbre explícita, integrando dimensiones de verdad, indeterminación y falsedad que no son capturadas por modelos deterministas clásicos.

En este sentido, los enfoques configuracionales presentan ventajas interpretativas frente a modelos de aprendizaje automático complejos, lo cual coincide con los principios del campo emergente de Explainable Artificial Intelligence (XAI), orientado a mejorar la transparencia e interpretabilidad de los modelos computacionales utilizados en procesos de toma de decisiones, cuyo objetivo es mejorar la transparencia e interpretabilidad de los modelos computacionales utilizados en procesos de toma de decisiones, cuyo objetivo es mejorar la transparencia y comprensión de los modelos computacionales utilizados en la toma de decisiones. En este sentido, el aprendizaje automático fue utilizado como mecanismo complementario de validación y no como sustituto del análisis configuracional, reforzando la consistencia empírica sin desplazar el enfoque causal estructural.

En relación con la validación predictiva, el desempeño del modelo SVM evidenció una exactitud aproximada del 89.3 %, lo cual indica consistencia en la identificación de patrones asociados a la violencia. No obstante, la matriz de confusión reveló un desbalance de clases, fenómeno común en estudios sociales donde ciertas categorías concentran mayor frecuencia. Este comportamiento coincide con observaciones previas en investigaciones empíricas sobre violencia juvenil, donde la distribución desigual de casos influye en las métricas de desempeño.

En conjunto, los resultados obtenidos no solo corroboran los hallazgos previos sobre la naturaleza multicausal de la violencia, sino que también aportan una contribución metodológica al integrar el enfoque configuracional con lógica neutrosófica y validación mediante aprendizaje automático, fortaleciendo la toma de decisiones basada en evidencia en contextos educativos y sociales.

Si bien el modelo propuesto mostró consistencia en la identificación de configuraciones causales relevantes, el estudio presenta ciertas limitaciones. En primer lugar, el conjunto de datos utilizado corresponde a un contexto geográfico específico, lo que podría influir en la generalización de los resultados a otras realidades socioculturales. En segundo lugar, la asignación automática de valores neutrosóficos se basó en reglas de normalización lineal, las cuales podrían ampliarse en investigaciones futuras mediante funciones de pertenencia no lineales o enfoques híbridos.

En términos de escalabilidad, la arquitectura desarrollada permite el procesamiento de volúmenes mayores de datos; sin embargo, futuras investigaciones podrían evaluar el desempeño del sistema en escenarios de big data y entornos multiinstitucionales. Asimismo, se propone explorar la integración de técnicas de aprendizaje profundo para contrastar configuraciones identificadas bajo diferentes paradigmas computacionales.

V. Conclusión

El estudio permitió identificar configuraciones causales asociadas a la violencia mediante la aplicación del enfoque configuracional neutrosófico propuesto.

Desde una perspectiva científica, el estudio aporta una contribución metodológica al formalizar el análisis configuracional bajo representación neutrosófica, una contribución computacional al implementar un motor configuracional reproducible para la identificación de patrones causales, y una contribución empírica al validar el enfoque propuesto mediante su aplicación a un conjunto de datos reales de gran escala.

En la fase de carga y validación se trabajó con un conjunto de datos sociales y educativos de gran escala, los cuales fueron divididos en un conjunto de entrenamiento (80 %) y un conjunto de prueba (20 %). Esta organización permitió evaluar el comportamiento del modelo bajo condiciones controladas y analizar su consistencia predictiva. Durante la transformación neutrosófica, las variables seleccionadas fueron representadas mediante grados de verdad, indeterminación y falsedad, lo que permitió modelar la incertidumbre inherente a los datos sociales. Esta etapa resultó clave para evitar interpretaciones binarias y capturar escenarios intermedios presentes en los indicadores analizados.

Desde una perspectiva científica, el presente estudio contribuye al fortalecimiento del Análisis Cualitativo Comparativo al incorporar formalmente la modelación neutrosófica como mecanismo para representar incertidumbre estructural en fenómenos sociales complejos. Esta integración amplía el marco tradicional del QCA al permitir la inclusión explícita de indeterminación en la evaluación configuracional, aportando una extensión metodológica aplicable a estudios multicausales en ciencias sociales.

En la aplicación del algoritmo SQA se identificaron configuraciones recurrentes donde el nivel educativo mostró relevancia cuando fue evaluado en combinación con otros factores. En contraste, variables como sexo y edad presentaron baja influencia cuando se analizaron de manera aislada, lo que confirma que el fenómeno estudiado no responde a una única causa, sino a estructuras configuracionales.

La validación mediante modelos de clasificación supervisada evidenció la consistencia predictiva del enfoque propuesto, alcanzando una exactitud aproximada del 89.3 % y un F1-score cercano a 0.86.

Los resultados evidencian que la integración del análisis configuracional neutrosófico con modelos de aprendizaje automático permite abordar fenómenos sociales caracterizados por alta incertidumbre desde una perspectiva estructural y predictiva. Mientras el análisis configuracional identifica combinaciones causales interpretables asociadas al fenómeno estudiado, los modelos de aprendizaje automático permiten validar empíricamente la capacidad predictiva de dichas configuraciones. Esta integración metodológica contribuye al desarrollo de herramientas analíticas más robustas para el estudio de fenómenos sociales caracterizados por alta incertidumbre.

VI. Agradecimiento

Los autores agradecen a los docentes de la carrera de Ingeniería en Software de la Universidad de Guayaquil por su orientación y apoyo en el desarrollo del presente trabajo.

VII. Referencias

- [1] L. A. Zadeh, “**Fuzzy sets,**” **Information and Control**, vol. 8, pp. 338–353, 1965, doi:10.1016/S0019-9958(65)90241-X.
- [2] K. T. Atanassov, “**Intuitionistic fuzzy sets,**” **Fuzzy Sets and Systems**, vol. 20, no. 1, pp. 87–96, 1986, doi:10.1016/S0165-0114(86)80034-3.
- [3] C. C. Ragin, **Redesigning Social Inquiry: Fuzzy Sets and Beyond**. Chicago, IL, USA: University of Chicago Press, 2008.
- [4] T. Hastie, R. Tibshirani, and J. Friedman, **The Elements of Statistical Learning, 2nd ed.** New York, NY, USA: Springer, 2009, doi:10.1007/978-0-387-84858-7.
- [5] J. Han, M. Kamber, and J. Pei, **Data Mining: Concepts and Techniques, 3rd ed.** Burlington, MA, USA: Morgan Kaufmann, 2011.
- [6] C. Cortes and V. Vapnik, “**Support-vector networks,**” **Machine Learning**, vol. 20, no. 3, pp. 273–297, 1995, doi:10.1007/BF00994018.
- [7] L. Breiman, “**Random forests,**” **Machine Learning**, vol. 45, pp. 5–32, 2001, doi:10.1023/A:1010933404324.



- [8] F. Smarandache, A Unifying Field in Logics: **Neutrosophic Logic, Neutrosophic Set, Neutrosophic Probability and Statistics**. Rehoboth, NM, USA: American Research Press, 2005.
- [9] F. Smarandache, “**Neutrosophic Sets and Systems**,” Neutrosophic Sets and Systems, vol. 8, pp. 1–8, 2013.
- [10] J. L. Salmeron and F. Smarandache, “**Artificial intelligence and neutrosophic logic**,” Neutrosophic Sets and Systems, vol. 20, pp. 1–8, 2018.
- [11] S. Pramanik, “**Neutrosophic sets and decision making**,” Neutrosophic Sets and Systems, vol. 12, pp. 16–22, 2016.
- [12] R. Morales and L. S., “**Factores sociales asociados a la violencia juvenil**,” Revista Iberoamericana de Psicología, vol. 14, no. 3, pp. 89–104, 2021.
- [13] J. Pérez and A. G., “**Análisis de la violencia escolar en contextos urbanos**,” Informe de investigación, 2022.
- [14] UNICEF, “**Violence Against Children**,” 2024. [Online]. Available: <https://data.unicef.org/topic/child-protection/violence/>
- [15] Human Rights Watch, “**Like Patchwork: Ecuador’s slow progress tackling and preventing school-related sexual violence**,” 2024.
- [16] S. Broumi, F. Smarandache, and M. Ali, “**Neutrosophic sets and systems: A structured survey**,” Information Sciences, vol. 615, pp. 154–183, 2023. <https://doi.org/10.1016/j.ins.2022.10.024>.
- [17] T. Akram, M. Saeed, and F. Smarandache, “**Recent advances in neutrosophic logic and applications**,” IEEE Access, vol. 10, pp. 105000–105020, 2022. <https://doi.org/10.1109/ACCESS.2022.3211433>