

FACULTAD DE INGENIERÍA

PROGRAMA DE INGENIERÍA DE SISTEMAS

Departamento de Tecnologías de Información y  
Telecomunicaciones

PROYECTO DE GRADO

Implementación de un modelo predictivo de feminicidios: Un enfoque ético de  
inteligencia artificial para abordar la desigualdad social y de género en  
Colombia

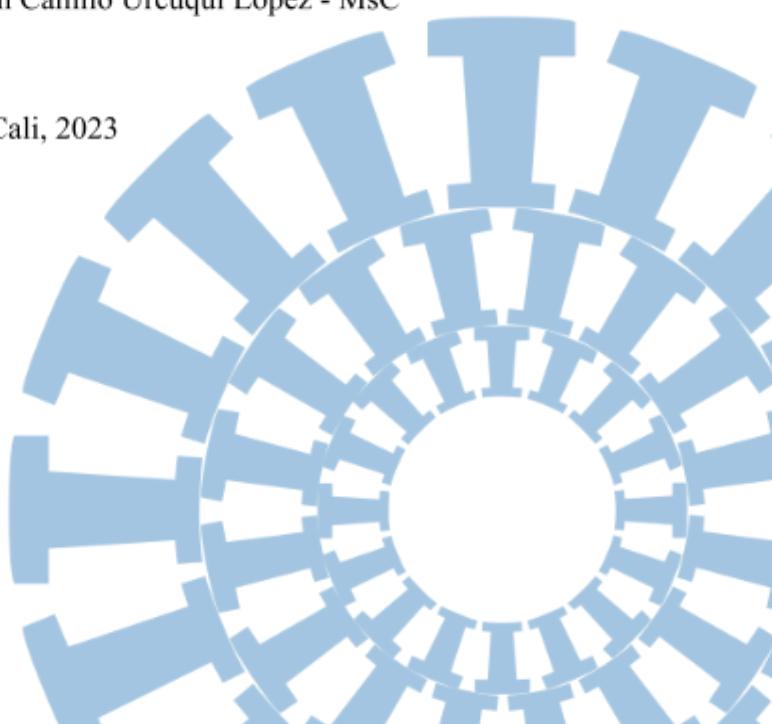
INTEGRANTES

Juan José Restrepo Bonilla - Sergio Alejandro Cabrera Nieto

DIRIGIDO POR

Juan Camilo Tobar Nuñez. Christian Camilo Urcuqui López - MsC

Santiago de Cali, 2023



Resumen	3
Abstract	3
Lista de acrónimos	4
Índice de figuras	5
Índice de tablas	6
1. Motivación y Antecedentes	6
1.1. Motivación	6
1.2. Antecedentes	7
2. Descripción del problema	8
2.1. Identificación del problema	8
2.2. Formulación del problema	8
3. Objetivos del proyecto	9
3.1. Objetivo general	9
3.2. Objetivos específicos	9
4. Marco teórico y estado del arte	9
4.1. Generalidades	9
4.1.1. Concepto de feminicidios y su prevalencia en Colombia	9
4.1.2. El Observatorio de Feminicidios de Colombia y su relevancia	11
4.2. Inteligencia Artificial y Aprendizaje Profundo	15
4.2.1. Definición de Inteligencia Artificial	15
4.2.2. Definición de Aprendizaje Profundo	16
4.3. Redes Neuronales Convolucionales y Long Short-Term Memory	17
4.3.1. Definición de Redes Neuronales Convolucionales	17
4.3.2. Definición de Long Short-Term Memory (LSTM)	19
4.3.3. ConvLSTM: Combinación de CNN y LSTM	19
4.4. Redes Adversarias Generativas	20
4.4.1. Definición de Redes Adversarias Generativas	20
4.4.2. Uso de cGANs para la predicción de datos temporales	22
4.5. Tensorflow	24
4.5.1. Generalidades sobre TensorFlow	24
4.5.2. Implementación de ConvLSTM y cGAN en TensorFlow	25
4.6. Marco Ético para la Inteligencia Artificial en Colombia	25
4.6.1. Objetivo y alcance	25
4.6.2. Principios éticos	26
4.6.3. Herramientas para la implementación	34
4.7. Estado del Arte	42
4.7.1. Aplicaciones de IA para la predicción de crímenes en el sector académico	42
4.7.2. Uso de IA para la predicción de crímenes en el sector público	43
5. Desarrollo del proyecto	46
5.1. Metodología	46
5.2. Fases de desarrollo del proyecto	47

5.2.1. Obtención de datos	47
5.2.2. Análisis y preparación de datos para el modelado	48
5.2.3. Fase de diseño	49
5.2.3.1. Diseño del modelo ConvLSTM-cGAN	49
5.2.3.2. Elección de parámetros y estructura de la red	50
5.2.4. Fase de implementación	53
5.2.4.1. Codificación y entrenamiento del modelo en TensorFlow	53
5.2.4.2. Iteraciones y mejoras del modelo	56
5.2.5. Fase de validación	58
5.2.5.1. Evaluación del modelo y ajustes necesarios	58
5.2.6. Primera fase de ajustes al diseño, implementación y validación al modelo planteado	59
5.2.7. Segunda fase de ajustes al diseño, implementación y validación al modelo planteado	69
5.3. Aportes:	83
5.4. Entregables:	84
5.5. Conclusiones:	85
6. Referencias bibliográficas	87
7. Anexos	91

## Resumen

En el presente proyecto de grado queremos abordar el delito de feminicidio en Colombia, que se caracteriza por ser una expresión extrema de desigualdad social y de género, con el fin de aplicar los conocimientos, con una de las recientes herramientas en inteligencia artificial. Se analizan en detalle los conceptos y prevalencias del feminicidio en nuestro país, y su evolución dentro de la legislación colombiana. Se proporciona un marco teórico completo sobre IA y aprendizaje profundo, enfocándose en Redes Neuronales Convolucionales (CNN) y Long Short-Term Memory (LSTM), con especial énfasis en Redes Generativas adversarias (GANs) y su papel en la predicción de series temporales de datos. Se describe el uso de Tensorflow para la implementación de ConvLSTM y cGANS.

Además, se analiza el marco ético para la IA en Colombia, reforzando la importancia de la ética en la aplicación de estas tecnologías. A lo largo del proyecto, se explora el estado del arte en el uso de IA para la predicción de delitos y se detallan las metodologías de desarrollo del proyecto, desde la adquisición de datos, hasta la validación de modelos. El objetivo principal es el desarrollo e implementación de un modelo predictivo basado en IA, para prevenir el feminicidio y contribuir a la reducción de la desigualdad social y de género en Colombia.

## Abstract

This degree project addresses femicide in Colombia, an extreme expression of social and gender inequality, applying knowledge in artificial intelligence. It analyzes in detail the concept and prevalence of femicide in our country, and its evolution within the Colombian legislation. A complete theoretical framework on AI and deep learning is provided, focusing on Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM), with special emphasis on Generative Adversarial Networks (GANs) and their role in the prediction of time series data. The use of Tensorflow for the implementation of ConvLSTM and cGANS is described.

In addition, the ethical framework for AI in Colombia is analyzed, reinforcing the importance of ethics in the application of these technologies. Throughout the project, the state of the art in the use of AI for crime prediction is explored and the project development

methodology is detailed, from data acquisition to model validation. The main objective is the development and implementation of an AI-based predictive model to prevent femicide and contribute to the reduction of social and gender inequality in Colombia.

## Lista de acrónimos

Símbolo	Descripción
AI	Inteligencia Artificial
ANN	Red Neuronal Artificial
CNN	Red Neuronal Convolucionar
LSTM	Long Short-Term Memory
RNN	Red Neuronal Recurrente
GAN	Red Generativa Antagonista
cGAN	Red Generativa Antagonista Condicional
ConvLSTM	Red Neuronal Convolucionar de Memoria a Largo Plazo
MAE	Error Absoluto Medio
MSE	Error Cuadrático Medio
RMSE	Raíz del Error Cuadrático Medio
DL	Aprendizaje Profundo
ML	Aprendizaje Automático
SGD	Descenso del Gradiente Estocástico
ReLU	Unidad Lineal Rectificada
Sigmoid	Función Sigmoide

Softmax	Función Softmax
Tanh	Función Tangente Hiperbólica
Adam	Optimizador Adam
EPOCH	Una iteración completa a través del conjunto de datos de entrenamiento
BATCH	Subconjunto del conjunto de datos de entrenamiento
Conv1D	Capas convolucionales unidimensionales
tándem	Se refiere a la práctica de entrenar dos modelos juntos, uno después del otro

## Índice de figuras

Figura	Título	Página
Figura 1	Informe anual 2020: “No somos un dia”	11
Figura 2	Panoramas feminicidios durante en el 2020	12
Figura 3	Lugar donde ocurrieron los hechos de feminicidio en Colombia	13
Figura 4	Relación que se tiene con la mujer	13
Figura 5	Diagrama de Capas de RNC 1 (Andrej Karpathy, n.d.)	17

Figura 6	Diagrama de Capas de RNC 2 (Andrej Karpathy, n.d.)	17
Figura 7	Diagrama de GAN (Oli N., 2018)	21
Figura 8	Diagrama Condicionamiento por concatenación (Mirza & Osindero, 2014)	22
Figura 9	Función de minimax (Mirza & Osindero, 2014)	23
Figura 10	(WWW Foundation, 2018)	39
Figura 11	Diagrama de riesgos (IMDA PDPC, 2020)	41
Figura 12	(Plan Internacional, AIME, 2021)	42

## Índice de tablas

Tabla	Título	Página
1	Tabla de resultados de modelos para multivariables de Forecasting	70

## 1. Motivación y Antecedentes

### 1.1. Motivación

Colombia, al igual que muchos países alrededor del mundo, ha sido testigo y partícipe con la afectación en un alto grado de incremento en la incidencia de la violencia de género,

particularmente el feminicidio. Este delito, que es la forma más extrema de violencia contra la mujer, ha generado una profunda preocupación a nivel nacional e internacional. Sin embargo, la magnitud y gravedad de esta problemática no han sido suficientemente contrarrestadas con políticas y estrategias efectivas para su prevención.

En este contexto, surge la motivación para explorar nuevos enfoques y herramientas que puedan contribuir a enfrentar este desafío. Entre estas, la inteligencia artificial (IA) aparece como una prometedora aliada, con su capacidad para analizar grandes volúmenes de datos y detectar patrones que podrían pasar desapercibidos para el análisis humano. A través de la aplicación de técnicas avanzadas de IA, se espera poder desarrollar modelos predictivos que ayuden a prevenir el delito de feminicidio en Colombia, contribuyendo a la protección de las mujeres y a la promoción para una sociedad Colombiana más equitativa y segura.

## 1.2. Antecedentes

El uso de la inteligencia artificial en la prevención del crimen es un campo de estudio que ha venido abriéndose espacio en los últimos años. Diversas investigaciones han demostrado el potencial de la IA para identificar tendencias y patrones en la comisión de delitos, y para prever situaciones de riesgo. Sin embargo, su aplicación específica para la prevención del feminicidio en Colombia es aún incipiente.

La legislación colombiana ha reconocido el feminicidio como un delito autónomo desde la implementación de la Ley Rosa Elvira Cely (Ley 1761 de 2015), pero las estadísticas muestran que se trata de un problema que persiste y se agrava. De hecho, Colombia ocupa uno de los primeros lugares en América Latina en cuanto a la prevalencia de este delito.

Al mismo tiempo, Muchas organizaciones que velan o luchan por la equidad de género, han venido liderando procesos que mitiguen los riesgos de la comisión del delito de Feminicidio que se ha acentuado en las clases sociales populares en Colombia, donde podemos coadyuvar a encontrar salidas con el aporte que ofrece la IA. El país ha mostrado interés en adoptar esta tecnología para enfrentar diversos desafíos en áreas como la salud, la educación, el medio ambiente y la seguridad. Aun así, la aplicación de la IA para la prevención del delito de feminicidio aún no ha sido explorada en profundidad, lo que indica que existe un importante espacio para la investigación y desarrollo en este campo.

El presente trabajo pretende llenar este vacío y proporcionar una base sólida para futuras investigaciones y aplicaciones prácticas en esta área de crucial importancia.

## 2. Descripción del problema

### 2.1. Identificación del problema

El feminicidio en Colombia es un delito que ha alcanzado proporciones alarmantes y requiere medidas efectivas para su prevención y mitigación. Diversos factores, como la violencia doméstica, la impunidad y la discriminación de género, contribuyen a la persistencia de este problema. Sin embargo, existe un vacío en cuanto a las estrategias de prevención basadas en el uso de tecnologías emergentes, como la inteligencia artificial (IA).

La IA tiene el potencial de proporcionar herramientas útiles para la detección temprana de situaciones de riesgo y para la identificación de patrones relacionados con el delito de feminicidio. No obstante, la implementación de la IA en este ámbito presenta desafíos éticos y técnicos, procurando cerrar la brecha de sesgos y la protección de la privacidad de los datos.

### 2.2. Formulación del problema

La pregunta central de este proyecto de grado es la siguiente: ¿Cómo puede la inteligencia artificial contribuir a la prevención y mitigación del delito de feminicidio en Colombia, de una manera ética y respetuosa con los derechos de las personas? Para responder a esta pregunta, será necesario explorar y evaluar distintas técnicas de IA, como el aprendizaje automático, en relación con su capacidad para predecir situaciones de riesgo y para apoyar la toma de decisiones en el ámbito de la prevención del delito de feminicidio.

Al mismo tiempo, será esencial considerar los desafíos éticos asociados con el uso de la IA, y desarrollar estrategias para minimizar los sesgos, garantizar la transparencia y proteger la privacidad de los datos.

### **3. Objetivos del proyecto**

#### **3.1. Objetivo general**

Diseñar e implementar un modelo de inteligencia artificial para contribuir a la prevención y mitigación del delito de feminicidio en Colombia, considerando las implicaciones éticas y los desafíos técnicos asociados con el uso de esta tecnología en el ámbito de la seguridad pública.

#### **3.2. Objetivos específicos**

- Recopilar y analizar datos relevantes sobre el delito de feminicidio en Colombia, incluyendo las características de las víctimas, los perpetradores y las circunstancias de los delitos.
- Identificar las técnicas de inteligencia artificial más adecuadas para la detección temprana de situaciones de riesgo y la identificación de patrones relacionados con el delito de feminicidio.
- Diseñar un modelo de inteligencia artificial para la prevención del feminicidio, basado en el análisis de los datos y en la selección de las técnicas de IA.
- Evaluar la eficacia y la ética del modelo de IA, considerando criterios como la minimización de sesgos, la transparencia y la protección de la privacidad de los datos.
- Proporcionar recomendaciones para la implementación y el uso ético de la inteligencia artificial en la prevención del delito de feminicidio en Colombia.

### **4. Marco teórico y estado del arte**

#### **4.1. Generalidades**

##### **4.1.1. Concepto de feminicidios y su prevalencia en Colombia**

En su trabajo sobre el feminicidio en Colombia, Ramírez Arce (2018) explora “...del proceso de evolución, que ha tenido el concepto de feminicidio a lo largo de los años en

nuestro país y de esta manera determinar mediante qué criterios se puede catalogar una conducta de asesinato contra una mujer como feminicidio y no como homicidio dentro de la legislación penal colombiana. Además de evidenciar la forma en como el feminicidio fue adoptado a lo largo del tiempo por parte de los organismos internacionales de derechos humanos y finalmente repercute en nuestro país con la implementación de nuevas normas al respecto convirtiéndolo como delito autónomo en nuestra legislación... ...y la manera en que dicho tipo penal entró a regir en el actual Código Penal, Ley 599 de 2000, mediante su adecuación inicial como agravante del Homicidio, expuesto por la ley 1257 de 2008 y finalmente como delito autónomo con la Ley 1761 de 2015 o comúnmente conocida Ley Rosa Elvira Cely".

Por otro lado tenemos dentro de este trabajo de investigación de Bernal Aldana (2020) argumenta que: "... Por otro lado, algunas investigaciones se han centrado en describir las consecuencias de la violencia contra las mujeres las cuales generan daños relevantes de su integridad. Por ende, un intento de feminicidio que, aunque como tal hasta ahora se está castigando y haciendo un poco más visible, el intento del mismo está aún más olvidado; esto implicaría reconocer que se requieren reformas muy importantes en la justicia. En cuanto a las cifras, se ha evidenciado que los mayores incrementos se registran en Chipre (277,13%), y Colombia (51,28%) y de acuerdo con las tasas de prevalencia de los feminicidios cometidos por cualquier agresor en el 2006, Colombia ocupa el tercer lugar en América Latina. Por ende". La investigación: ¿Será que a las mujeres nos matan porque nos aman?, señala que "La información analizada del período 2002-2009 acerca de la violencia en pareja, la violencia sexual, las lesiones personales, los homicidios perpetrados 4 en contra de las mujeres y los feminicidios, permite constatar el continuum de las violencias en contra de las mujeres y el feminicidio como su eslabón final" (Ley N° 107, 2013).

Ahora bien, en el Instituto Nacional de Medicina Legal de Colombia, en su Informe "Forensis 2010" describe que para el año 2010, las cifras arrojaron que 1.444 mujeres fueron asesinadas; en comparación con este, en el informe "Forensis 2018" se describe que para el año 2018, se desarrollaron 49.669 peritaciones en el contexto de la violencia de pareja, siendo el hombre, el principal presunto agresor; según esto y en relación al feminicidio, de los 404 casos de asesinatos de mujeres registrados en el año 2018, 78 fueron catalogados bajo la circunstancia feminicidio.".

En este concepto histórico, la identificación de la problemática, su estadística e implementación dentro de las normas penales colombianas, se hace relevante el aporte de las nuevas tecnologías de la información como una herramienta de prevención, mitigación e identificación de los posibles causales de esta conducta que genera el hombre contra la mujer, haciendo relevante el desarrollo y aplicación de la inteligencia artificial, que apuntan en buscar que estas cifras se reduzcan drásticamente y lo más importante, se logren salvar vidas de muchas colombianas con el aporte de esta herramienta digital, que estamos desarrollando para que las autoridades del orden municipal y departamental en los diferentes territorios de Colombia, al implementarla se cumpla el objetivo primordial que es la prevención y no consumación de este terrible delito que se ha ido incrementando a través de los años.

#### 4.1.2. El Observatorio de Feminicidios de Colombia y su relevancia

Según la Fundación Feminicidios Colombia (2020) : “Se identificó que la mayoría de feminicidios tentados y consumados se concentran en mujeres con edades productivas a nivel económico y social, la mayoría de ellas trabajaban y ejercían trabajos domésticos (19 a 45 años: 149 víctimas). Los hombres con una cultura machista arraigada presentan una tendencia marcada a matar mujeres en su mayor fase de productividad lo cual, se da como rechazo y castigo a la autonomía económica de aquellas que conlleva muchas veces a la decisión de terminar vínculos afectivos. Es así como, cuando finalmente toman la decisión de terminar sus relaciones vienen ataques letales ante la resistencia de retomar la relación, también se presentan cuando las mujeres deciden iniciar otras relaciones, indistintamente del tiempo transcurrido entre la finalización de una y el inicio de otra. Lo anterior, indica que no es verdad que los programas de empoderamiento económico de mujeres garantizan per se la disminución de violencias machistas, ya que aquellas no son la consecuencia, -como se ha querido afirmar durante años-, de dependencias emocionales y económicas de mujeres hacia los hombres, sino de la ausencia de reconocimiento de los segundos respecto de las decisiones de las primeras y la imposición de roles que al romperse, dan pie al castigo mortal, por no ser, aquello que la sociedad ha denominado “buenas mujeres”. Esto explica el hecho de que la mayoría de los feminicidios tentados y consumados sean de tipo íntimo (perpetrados por pareja o expareja).

**Edades de las víctimas:** 19 – 25años 23,3% Mujeres jóvenes, 26 – 36 años 27,9%, 37 – 45 años 26,7%, 46 – 60 años 8,4 %, + 60 años 2,9 %, n/r 5,5 % (Se extracto el rango de edad y los porcentajes más altos donde se han cometido los feminicidios). ”.

A continuación extractamos varios de los apartes del informe, que nos va a permitir contextualizar el panorama en el territorio nacional donde se evidencia la incidencia en rango de edades, tipo de relación, orden territorial y la concentración de este delito:

Figura 1

Probabilidades de ser víctimas de feminicidios en los propios hogares



*Nota.* Probabilidades de ser víctimas de feminicidios en los propios hogares.

Tomado de "No somos un día: Informe Anual 2020" por la Fundación Feminicidios Colombia, 2020, p. 5 (<https://www.feminicidioscolombia.org>).

Figura 2

Panoramas feminicidios durante en el 2020



\* Los casos en verificación corresponden a aquellos en los que hubo una muerte violenta de una mujer sin que se conozcan las circunstancias que permiten catalogarlos como un feminicidio.

Nota. Panoramas de feminicidios durante el 2020. Tomado de "No somos un día: Informe Anual 2020" por la Fundación Feminicidios Colombia, 2020, p. 4 (<https://www.feminicidioscolombia.org>).

Figura 3

Lugar donde ocurrieron los hechos de feminicidio en Colombia



Nota. Lugar donde ocurrieron los hechos de feminicidio en Colombia. Tomado de "No somos un día: Informe Anual 2020" por la Fundación Feminicidios Colombia, 2020, p. 4 (<https://www.feminicidioscolombia.org>).

Figura 4

Relación que se tiene con la mujer

**Relación con el victimario**

**En el 43.5% de los casos los victimarios eran parejas y en el 21,1% exparejas.**

**También se presentan casos con victimarios conocidos, familiares, incluyendo padres y padrastrós.**



Los cuales ocurren en la mayoría en el ámbito privado, primordialmente en las viviendas, siendo las armas cortopunzantes las de mayor uso ante la facilidad del acceso a ellas en ese ámbito, así como un decrecimiento en jornadas de desarme en el país.

De otro lado, se presentaron casos de feminicidio y tentativas en ámbitos públicos, siendo las vías públicas las más frecuentes. Ello está relacionado, con la presencia de modalidades de tipo sicarial para la comisión de delitos o la presencia de los feminicidas en los espacios que transitan las mujeres víctimas.

*Nota.* Relación que se tiene con la mujer. Tomado de "No somos un día: Informe Anual 2020" por la Fundación Feminicidios Colombia, 2020, p. 4 (<https://www.feminicidioscolombia.org>).

Según Hernández Severiche y Díaz-Rincón (2021) analizan el feminicidio dentro del contexto jurídico-penal colombiano y su incidencia en la sociedad: “En este ambiente de dominación por parte del hombre producto de la misma sociedad patriarcal que se ha encargado de posicionar los roles en esta misma de acuerdo con el género respectivo, da lugar a que se encasille a la mujer en respectivos oficios y se le concedan ciertos derechos lo que provoca que al momento en el que la mujer quiera obtener un rol dominante frente al hombre, este último se sienta intimidado optando por agredirla debido a su posición dominante en la sociedad.... .En este orden de ideas, el hecho de ser mujer en una sociedad discriminatoria, que somete a la mujer, la opprime a roles de subordinación, genera según Toledo (2012), un factor de riesgo y vulnerabilidad frente a la violencia contra estas. Es por ello por lo que factores como la impunidad y la tolerancia frente a casos de violencia basada en el género ha provocado la existencia de casos de feminicidio, pues estos tienen sus orígenes en la incapacidad por parte del Estado de prevenir y sancionar la violencia contra la mujer y, en consecuencia, conlleva a la impunidad como pan de cada día.” Contextualizando la mirada social, económica, laboral que ha querido romper con los orígenes patriarcales en nuestra sociedad colombiana, donde se ha querido enmascarar que la mujer debe apoyarse en su empoderamiento productivo y económico como herramienta para generar igualdad ante los hombres, generando por el contrario una dicotomía en la relación de hombre-mujer que han constituido una relación o un matrimonio, conllevando a que el hombre considere su

desplazamiento o la sumatoria de un rol más significativo en la sociedad generando a este, una amenaza que considera que vulnera su condición social como líder social afectivo o cabeza visible del núcleo familiar como es el hogar.

En este orden de ideas, es fundamental el desarrollo y la aplicación de tecnologías de la información y la inteligencia artificial que nos permita identificar o visualizar tempranamente lugares, acciones o posible consumación de un feminicidio por diferentes algoritmos que pretendemos desarrollar para atacar, mitigar, neutralizar y disminuir los índices o estadísticas donde se han perpetrado las muertes o secuelas de ataques por diferentes métodos, privando la utilización de armas blancas del hombre contra la mujer, por eso consideramos que el desarrollo de aplicaciones, metodologías, algoritmo y demás herramientas de las tecnologías de la información toman mayor relevancia para ser posible contar con esta importante herramienta que ponemos a disposición y consideración de la academia de los gobiernos de carácter local y regional y las demás autoridades que permite coadyuvar a bajar los índices de muerte por este delito y como resultado final, salvar vidas valiosas de las mujeres en la construcción de una mejor sociedad en los diferentes territorios que hacen y forman parte de la república de Colombia.

#### 4.2. Inteligencia Artificial y Aprendizaje Profundo

##### 4.2.1. Definición de Inteligencia Artificial

La Inteligencia Artificial (IA) es un campo de estudio en la informática que se centra en la creación de sistemas capaces de realizar tareas que normalmente requieren inteligencia humana. Según el portal de IBM (2021) define la Inteligencia Artificial como " En su forma más simple, la inteligencia artificial es un campo que combina la ciencia informática y los conjuntos de datos robustos para permitir la resolución de problemas. También abarca los subcampos del machine learning y el deep learning, que se mencionan frecuentemente junto con la inteligencia artificial. Estas disciplinas están conformadas por algoritmos de IA que buscan crear sistemas expertos que hagan predicciones o clasificaciones basadas en datos de entrada". Además, la IA incluye subcampos como el aprendizaje automático y el aprendizaje profundo, que utilizan algoritmos para crear sistemas expertos capaces de hacer predicciones o clasificaciones basadas en datos de entrada.

La IA puede clasificarse en dos tipos: IA débil y IA fuerte. La IA débil, que es el tipo de IA más comúnmente implementada hoy en día, está diseñada para llevar a cabo una tarea específica, como el reconocimiento de voz. Por otro lado, la IA fuerte es una forma de IA que tiene la capacidad de entender, razonar, aprender y comportarse de una manera indistinguible de la inteligencia humana.

#### 4.2.2. Definición de Aprendizaje Profundo

El Aprendizaje Profundo es una subárea de la Inteligencia Artificial que se centra en los algoritmos inspirados en la estructura y función del cerebro llamados redes neuronales artificiales. Como señala el portal de Amazon (2022): “El aprendizaje profundo es un método de la inteligencia artificial (IA) que enseña a las computadoras a procesar datos de una manera que se inspira en el cerebro humano. Los modelos de aprendizaje profundo son capaces de reconocer patrones complejos en imágenes, textos, sonidos y otros datos, a fin de generar información y predicciones precisas”. Es importante aclarar que para el presente trabajo sobre la definición del "profundo" en "aprendizaje profundo" se refiere a la cantidad de capas a través de las cuales los datos se transforman en estas redes neuronales. En una red profunda, hay más capas entre la entrada y la salida, lo que permite que el algoritmo modele relaciones más complejas.

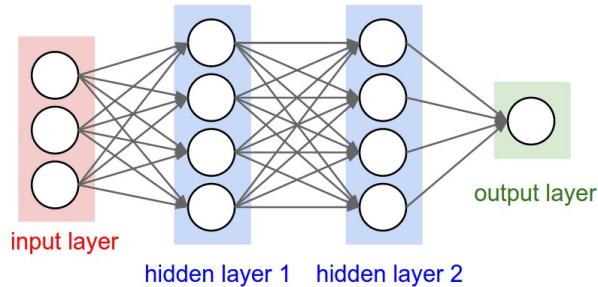
Además, es importante mencionar que el aprendizaje profundo se ha utilizado en una serie de aplicaciones revolucionarias. Desde la clasificación de imágenes hasta la traducción automática, pasando por el reconocimiento de voz, el aprendizaje profundo está transformando numerosos campos de la ciencia y la tecnología. Un punto clave en el éxito del aprendizaje profundo radica en su capacidad para procesar y aprender de grandes cantidades de datos. Cuando se alimenta con suficientes datos, un modelo de aprendizaje profundo puede aprender a realizar tareas muy complejas y precisas, superando a menudo a los humanos en ciertas tareas. Sin embargo, el aprendizaje profundo también presenta desafíos, como la necesidad de grandes cantidades de datos y poder de cómputo, la interpretación de los modelos y la posibilidad de sesgo en los datos de entrenamiento.

## 4.3. Redes Neuronales Convolucionales y Long Short-Term Memory

### 4.3.1. Definición de Redes Neuronales Convolucionales

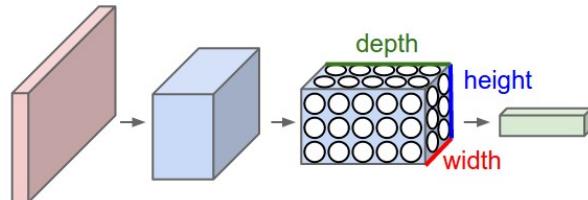
Para el presente trabajo, consideramos que lo que más se acerca para definir las Redes Neuronales Convolucionales (CNN, por sus siglas en inglés) pueden conceptualizarse como un mecanismo cerebral artificial que asiste a los sistemas computacionales en la interpretación y comprensión de las imágenes. Estas redes aplican un enfoque especializado para analizar las imágenes, permitiéndoles identificar elementos significativos dentro de las mismas. Este proceso asiste a la computadora en la identificación del contenido de la imagen y en la determinación de las acciones apropiadas basadas en esta información. (Andrej Karpathy, n.d.)

Figura 5  
Diagrama de Capas de RNC



*Nota.* Estructura de red neuronal densa. Cuenta con la primera capa de entrada, dos capas ocultas donde se encuentran el mayor número de neuronas y una capa de salida que cuenta con una sola neurona.

Figura 6  
Diagrama de capas de RNC



*Nota.* La capa de entrada roja tiene la imagen, por lo que su ancho y alto serían las dimensiones de la imagen, y la profundidad sería 3 (canales Rojo, Verde, Azul). Después de la imagen, la capa de entrada se somete a una serie de operaciones en las capas subsiguientes de la CNN.

De igual manera, las CNN están estructuradas en varias capas o bloques de construcción funcionales, cada uno con una tarea específica. Entre estas se encuentra la capa de convolución, que tiene como función principal ayudar a la red a aprender y reconocer patrones clave en las imágenes, como los bordes y otras características distintivas. Otra capa importante es la de agrupación (o pooling), que se encarga de simplificar y resumir la información obtenida en las capas anteriores, reduciendo la complejidad de la imagen pero conservando sus aspectos más relevantes. (Andrej Karpathy, n.d.)

Paralelamente, una ConvNet organiza sus neuronas en tres dimensiones (ancho, alto, profundidad), como se puede observar en una de sus capas. Cada capa de una ConvNet transforma el volumen de entrada 3D en un volumen de salida 3D de activaciones de neuronas. Como ejemplo, la capa de entrada, que puede ser representada en rojo, contiene la imagen. En esta capa, el ancho y la altura serían las dimensiones de la imagen, y la profundidad sería 3 (correspondiente a los canales Rojo, Verde, Azul). (Andrej Karpathy, n.d.)

Una ConvNet se compone de diversas capas. Cada capa posee una API simple: transforma un volumen de entrada 3D en un volumen de salida 3D con alguna función diferenciable que puede o no tener parámetros. (Andrej Karpathy, n.d.)

Finalmente, la red cuenta con una capa totalmente conectada, que recopila y sintetiza la información de todas las capas anteriores. Este componente de la red es fundamental, ya que integra todos los datos procesados por las capas previas, permitiendo la identificación final de la imagen y proporcionando una base para las decisiones y acciones del sistema. (Andrej Karpathy, n.d.)

#### 4.3.2. Definición de Long Short-Term Memory (LSTM)

De acuerdo al portal web colah github, el autor Olah (2015) define a las Redes Neuronales de Largo Corto Plazo (LSTM) son un tipo especial de Redes Neuronales Recurrentes (RNN) que tienen la habilidad de aprender dependencias a largo plazo. Las LSTM están diseñadas específicamente para evitar el problema de la desaparición del gradiente, un problema común en las RNN tradicionales.

En todas las RNN, la estructura se compone de una cadena de módulos de red neuronal repetitivos. Sin embargo, las LSTM tienen una estructura muy especial. En lugar de tener un solo estado de red neuronal, las LSTM tienen lo que se denomina celdas de memoria. Estas celdas de memoria se extienden a lo largo de las cadenas, con tres tipos de "puertas" que controlan el flujo de información hacia adentro y afuera de la celda de memoria (Olah, 2015).

Las "puertas" en las LSTM son mecanismos que permiten opcionalmente el paso de información. Se componen de una capa de red neuronal sigmoide y una operación de multiplicación puntual. Estas puertas son: la puerta de olvido, que decide qué información se va a eliminar de la celda de memoria; la puerta de entrada, que decide qué valores de la celda de estado se van a actualizar; y la puerta de salida, que decide cuál será la salida de la celda de estado. La salida será basada en el estado de la celda, pero será una versión filtrada (Olah, 2015).

Estas características hacen de las LSTM un modelo muy eficaz para tareas de aprendizaje secuencial, ya que puede retener información a largo plazo y decidir qué información es relevante para mantener o desechar (Olah, 2015).

#### 4.3.3. ConvLSTM: Combinación de CNN y LSTM

De acuerdo con el portal Medium, el autor Pokharna (2016) define las Redes Neuronales Convolucionales (CNN) se utilizan principalmente para el procesamiento de imágenes, en el cual cada imagen de entrada se trata como una matriz de píxeles. Las CNN son capaces de capturar con éxito las dependencias espaciales y temporales en una imagen a través de la aplicación de filtros relevantes.

La arquitectura juega un papel crucial para facilitar este tipo de aprendizaje. Las CNN constan de varias capas de neuronas. Cada una de estas capas aplica diferentes filtros y está seguida de otras capas como la ReLU (Unidad Lineal Rectificada) y la capa de pooling o de agrupamiento, y al final se utilizan capas completamente conectadas para la clasificación final. Este enfoque en capas reduce la cantidad de parámetros, haciendo que el modelo de red sea fácil de entrenar (Pokharna, 2016).

En cuanto a las ConvLSTM, son una extensión de la arquitectura de las LSTM para secuencias donde los datos de entrada son imágenes. Las ConvLSTM reemplazan las operaciones de productos matriciales de las LSTM por operaciones de convolución. Este cambio las hace más adecuadas para tareas donde los datos tienen una estructura de cuadrícula espacial, como es el caso de una imagen o un video. Así, las ConvLSTM son capaces de capturar tanto las dependencias espaciales (a través de convoluciones) como las temporales (a través de las LSTM) en los datos (Pokharna, 2016).

#### 4.4. Redes Adversarias Generativas

##### 4.4.1. Definición de Redes Adversarias Generativas

Las Redes Generativas Adversarias (GAN, por sus siglas en inglés) son algoritmos de aprendizaje automático que fueron propuestas por Ian GoodFellow y sus colegas en el 2014. Las GANs son esencialmente dos redes neuronales que compiten entre sí en un juego de suma cero (Goodfellow et al., 2014). Adicionalmente, existen otros tipos de GANs como lo son: DCGAN (Deep Convolutional GAN), cGAN (Conditional GAN), infoGAN (Information Maximizing GAN), etc. En la cual cada una cumple una necesidad específica, dependiendo del uso y de cómo esté implementada esta.

Las GANs constan de dos partes principales:

1. **Generador:** Esta red neuronal, toma un vector de ruido aleatorio y genera una muestra de datos a partir de ese vector. El objetivo principal del generador es producir datos que sean confusos o indetectables de los datos reales.
2. **Discriminador:** Esta red neuronal, toma un conjunto de datos, que pueden ser reales o pueden ser generados y los clasifica a cada uno de los datos si es real o es falso (es

decir, generado por el generador).

El juego de suma cero, mencionado anteriormente, se desarrolla de la siguiente manera: El generador trata de aumentar las posibilidades de que el discriminador se equivoque, mientras que el discriminador intenta clasificar correctamente los datos como reales o falsos. Esta dinámica competitiva entre ambas redes neuronales, permite que el generador aprenda a producir datos, cada vez más realistas a medida que el discriminador se vuelva más eficiente en la detección de datos falsos.

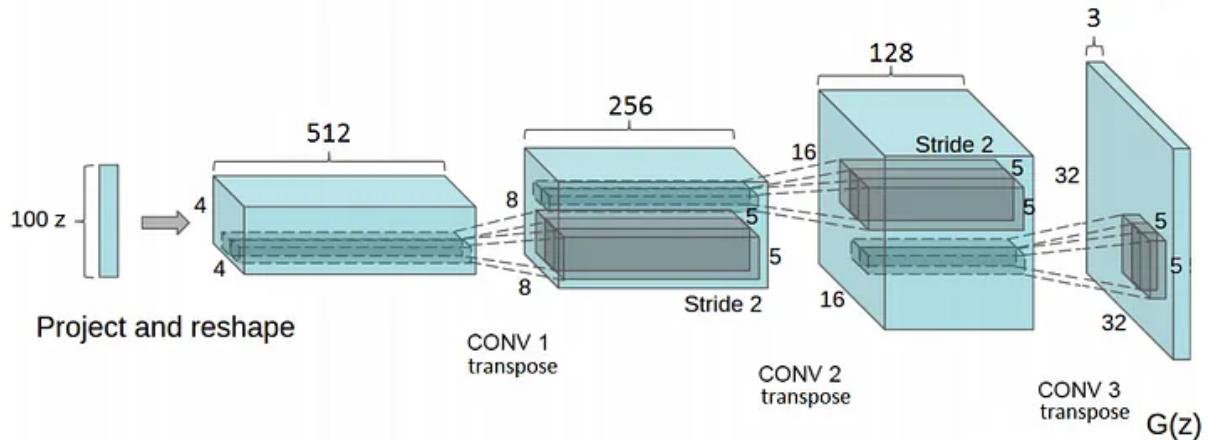
A continuación se presenta un resumen del juego, el cual continúa de la siguiente manera:

- El generador intenta maximizar la probabilidad de hacer que el discriminador confunda sus entradas como reales.
- Y el discriminador guía al generador para producir imágenes más realistas.

En el equilibrio perfecto, el generador capturará la distribución general de los datos de entrenamiento. Como resultado, el discriminador siempre estaría inseguro de si sus entradas son reales o no (Goodfellow et al., 2014). Adicionalmente a esto, las GANs se han podido utilizar en una gran variedad de aplicaciones y han tenido éxito, que incluyen la generación de imágenes realistas, la traducción de imágenes, la superresolución, el relleno de imagen y muchas más. Facilitando así muchas áreas ya sean de salud, de educación, entretenimiento, realidad virtual y en lo que nos estamos proponiendo a aportar en la prevención y mitigación del delito de feminicidio (Isola et al., 2017; Ledig et al., 2017; Iizuka et al., 2017).

Figura 7

Diagrama de GAN



*Nota.* Aquí se muestra la representación de una GAN y los procesos que esta conlleva

#### 4.4.2. Uso de cGANs para la predicción de datos temporales

Los datos temporales, comúnmente conocidos como datos de series de tiempo, son tipos de datos que se recopilan secuencialmente a lo largo del tiempo o periodo que se define. Como señala Na8 (2019), "una serie temporal es un conjunto de muestras tomadas a intervalos de tiempo regulares". En este contexto, los puntos de datos dentro de una serie de tiempo tienen asociadas unas marcas de tiempo, lo que se traduce en que tienen un orden específico y se registran en intervalos regulares o irregulares. Por ejemplo, las lecturas de temperatura que son tomadas a cada hora durante un día, las cotizaciones en la bolsa o las ventas registradas anualmente de un producto, todas ellas hacen parte del concepto de datos temporales.

En el contexto de los modelos de inteligencia artificial o machine learning, la interpretación y tratamiento de los datos temporales dependen del contexto y la naturaleza específica del problema que se está abordando, lo cual muchas veces, pueden ser muy desafiantes a la hora de su implementación, debido a sus dependencias temporales, que a menudo requieren de técnicas especiales de modelado, como las Redes Neuronales Recurrentes (RNN) o las redes generativas adversarias Condicionales (cGANs), para manejar correctamente estas dependencias.

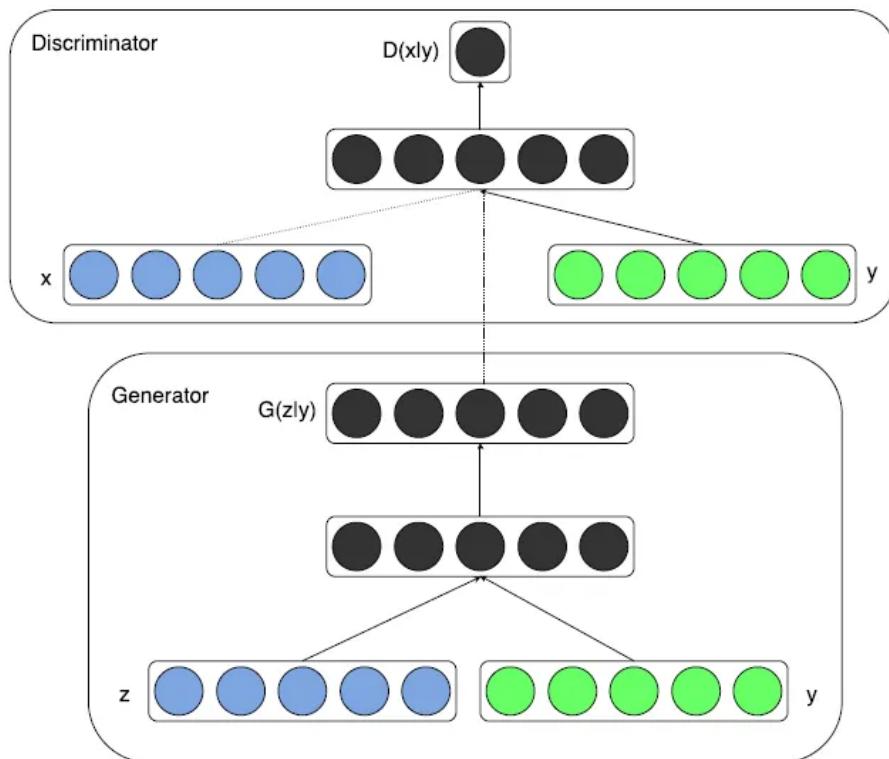
Además de su capacidad para modelar la distribución subyacente de los datos temporales, las cGANs tienen una ventaja particular, dado a que son capaces de generar múltiples posibles futuros. Esto puede ser útil, en situaciones o problemáticas donde exista la incertidumbre inherente en la evolución futura de una serie temporal. Sin embargo, como

cualquier otra técnica de modelado, el uso y la implementación de las cGANs para la predicción de datos temporales tiene sus desafíos. Estos pueden incluir la dificultad de seleccionar cual es la mejor arquitectura adecuada a la red neuronal, así como la necesidad significativa de datos de entrenamiento. En las siguientes secciones, profundizaremos en las técnicas de las cGANs y explicaremos su aplicabilidad y eficacia para la predicción de datos temporales en varios contextos.

A continuación se muestran varios gráficos sobre la estructura de red en una cGAN:

Figura 8

Diagrama Condicionamiento por concatenación



*Nota.* La estructura de una red adversaria condicional simple se muestra en la imagen superior. (Mirza & Osindero, 2014)

- Las redes generativas adversarias pueden extenderse a un modelo condicional si tanto el generador como el discriminador se condicionan en base a alguna información adicional  $y$ . (Mirza & Osindero, 2014)

- La variable  $y$  podría ser cualquier tipo de información auxiliar, como etiquetas de clase o datos de otras modalidades.
- La condición se realiza alimentando con  $y$  tanto al discriminador como al generador como una capa de entrada adicional (Mirza & Osindero, 2014).
- En el generador, el ruido de entrada anterior  $p_z(z)$  y  $y$  se combinan en una representación oculta conjunta.
- En el discriminador,  $x$  y  $y$  se presentan como entradas a una función discriminatoria.
- La función objetivo de un juego minimax de dos jugadores se convierte en la ecuación presentada (Mirza & Osindero, 2014).

Figura 9

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x|y)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z|y)))].$$

## 4.5. Tensorflow

### 4.5.1. Generalidades sobre TensorFlow

Tensorflow es una biblioteca de software de código abierto que tiene un uso en alto rendimiento para cómputo numérico, desarrollado por Google Brain Team (Abadi et al., 2015), su uso principalmente se deriva de aplicaciones de Machine Learning y Deep Learning debido a su gran capacidad para manejar gráfico de flujo de datos en múltiples CPUs o GPUs y por su gran variedad de herramientas de soporte.

El marco de trabajo de TensorFlow permite el desarrollo y entrenamiento de modelos de aprendizaje profundo utilizando una gama de redes neuronales, teniendo soporte para la distribución de cálculos a través de múltiples dispositivos para acelerar el procesamiento (Abadi et al., 2015), gracias a esto, TensorFlow proporciona múltiples niveles de abstracción, lo que permite la implementación de modelos tanto a nivel de operaciones matemáticas individuales como la utilización de Apis como “tf.keras”, que facilitan la construcción y el entrenamiento de modelos más robustos.

#### 4.5.2. Implementación de ConvLSTM y cGAN en TensorFlow

La implementación de la cGAN (Conditional Generative Adversarial Network) y la ConvLSTM (Long Short-Term Memory convolucional) en TensorFlow se facilita gracias a las bibliotecas y funciones de alto nivel proporcionadas por TensorFlow, específicamente Keras, la API de alto nivel para la construcción y el entrenamiento de modelos en TensorFlow (Goodfellow et al., 2014). La arquitectura de la cGAN consta de un discriminador y un generador. En la implementación elaborada, la clase “cGAN” hereda de “keras.model”, permitiendo la construcción de este modelo personalizado de cGAN. El método “Train\_step” define el proceso de entrenamiento durante una época, generando primero secuencias falsas a partir de vectores latentes, que se combinan con secuencias reales y se alimentan al discriminador para el entrenamiento.

Las redes ConvLSTM son adecuadas para el manejo de datos secuenciales unidimensionales, como es el caso de este proyecto. El generador y el discriminador, que se construyen en las funciones “build\_cgan\_generator” y “build\_cgan\_discriminator”, son ejemplos de estas redes. En ambos casos, se utiliza la función “LeakyReLU” en sus capas ocultas, y se combina la entrada de los datos y las etiquetas concatenándolas a lo largo del último eje.

### 4.6. Marco Ético para la Inteligencia Artificial en Colombia

#### 4.6.1. Objetivo y alcance

El "Marco Ético para el Desarrollo y la Implementación de la IA en Colombia" es una guía diseñada para proporcionar sugerencias y recomendaciones a las Entidades Públicas con respecto a la formulación y gestión de proyectos que implementen la Inteligencia Artificial (IA) (OECD, 2021). Cabe destacar que esta guía no es de cumplimiento obligatorio, sino que se presenta como un conjunto de recomendaciones basadas en las prácticas internacionales. En el informe de la OCDE de 2021, se reconoció como un mecanismo de 'soft law' debido a su naturaleza y la adopción que está recibiendo desde el Gobierno Nacional.(Guío Español et al., 2021).

El objetivo final es que las entidades estatales sigan estos principios en el desarrollo de los sistemas de IA y compartan públicamente cualquier información relacionada con este documento. Este proceso está destinado a generar transparencia y confianza en los sistemas por parte de los ciudadanos (OECD, 2021).

En cuanto a su alcance, este marco sugiere la aplicación de principios éticos para proyectos que utilizan IA y proporciona herramientas para la implementación de estos principios y algunas recomendaciones aplicables. Está dirigido a líderes estratégicos de TI, profesionales de TI, gerentes de proyectos internos y externos, profesionales de planificación, oficinas de gestión de proyectos (PMO) y profesionales responsables de la implementación de la Política de Gobierno Digital en las Entidades Públicas del Estado Colombiano.(Guío Español et al., 2021).

El documento indica que la implementación de este Marco Ético será una prioridad para las entidades del sector público. Sin embargo, esto no significa que no pueda ser implementado por entidades de otros sectores que estén desarrollando o implementando sistemas de IA. El Gobierno Nacional apoya proactivamente todas las iniciativas que busquen implementar estos principios en sectores fuera del Gobierno, así como a nivel regional y en entidades departamentales y municipales (OECD, 2021).

#### 4.6.2. Principios éticos

**Transparencia y explicabilidad:** Este principio se enfoca en proporcionar información abierta, significativa y comprensible sobre el diseño, funcionamiento e impacto de los sistemas de inteligencia artificial (IA). La información debe ser de fácil acceso y comprensible, promoviendo la participación activa de los ciudadanos en todas las fases de los sistemas de IA. Aunque este principio prioriza la transparencia, no debe comprometer la confidencialidad del modelo de gestión y la innovación.(Guío Español et al., 2021).

- **Ética de los datos:** La transparencia y la explicabilidad en este contexto implican proporcionar información comprensible sobre el origen de la información utilizada en el diseño de la tecnología, las características de los datos y los propósitos para los que se utilizarán los datos, especialmente los datos personales. El objetivo es que las personas entiendan cómo se recopilan

y procesan los datos, y por qué, especialmente cuando se procesan datos personales.(Guío Español et al., 2021).

- **Ética de los algoritmos:** Este aspecto implica proporcionar información significativa que permita entender los objetivos iniciales y los resultados esperados y obtenidos del modelo. Es relevante proporcionar información a la ciudadanía sobre los insumos utilizados en el diseño de los sistemas y los posibles resultados (transparencia de entrada y salida). También es importante proporcionar información significativa y comprensible sobre los errores (falsos positivos o negativos) y los indicadores de rendimiento del algoritmo.(Guío Español et al., 2021).
  - **Ética de las prácticas:** En este contexto, se debe proporcionar información significativa, clara y comprensible sobre los roles de los equipos y las personas involucradas en el diseño, desarrollo e implementación de estos sistemas, los objetivos que persiguen y los códigos de conducta que han generado para desarrollar su trabajo. Estos códigos de conducta deben ser de acceso público. Además, se debe informar a los usuarios cuando interactúan con sistemas de IA y no haya humanos detrás de los contenidos o respuestas que se generan.(Guío Español et al., 2021).
- 

**Privacidad:** Este principio enfatiza el respeto por la intimidad y la esfera privada de las personas en el contexto de la inteligencia artificial (IA). Se prohíbe el uso de información no autorizada y el perfilamiento a través de esta tecnología.(Guío Español et al., 2021).

- **Ética de los datos:** La privacidad en este ámbito lleva a la necesidad de obtener autorización para el uso de información personal, a menos que se trate de datos públicos o que existan excepciones específicamente señaladas por la ley. Este principio obliga a desarrollar mecanismos para mejorar la calidad de los datos utilizados y para actualizar constantemente la información. Además, las personas afectadas deben tener la posibilidad de corregir información personal incorrecta o errónea que se utilice en el desarrollo u operación de estos sistemas.(Guío Español et al., 2021).

- **Ética de los algoritmos:** El diseño de los algoritmos debe respetar la intimidad de las personas, por lo que los criterios de decisión no deben basarse en características personales o propias de su esfera privada. Se debe limitar el uso de información personal y solo usar la necesaria para un adecuado funcionamiento del sistema. Los diseñadores deben evitar el desarrollo de tecnologías que faciliten el perfilamiento de las personas sin su conocimiento y consentimiento.(Guío Español et al., 2021).
  - **Ética de las prácticas:** En este contexto, deben existir procedimientos internos que promuevan buenas prácticas en el uso de la información y en las respuestas y explicaciones que se brindan a los usuarios impactados por estas tecnologías. Deben implementarse medidas de responsabilidad demostrada que permitan la implementación de herramientas de gestión de riesgo para la privacidad, así como mecanismos de análisis de impacto en privacidad. Los equipos de diseñadores y desarrolladores deben establecer criterios para identificar casos de perfilamiento, su impacto y cómo evitar resultados negativos.(Guío Español et al., 2021).
- 

**Control humano de las decisiones propias de un sistema de inteligencia artificial (Human-in-the-loop y Human-over-the-loop):** Este principio se refiere a la necesidad de que los humanos mantengan un control significativo sobre los sistemas de inteligencia artificial (IA), particularmente aquellos que tienen cierto grado de autonomía en la toma de decisiones.(Guío Español et al., 2021).

- **Ética de los datos:** En este ámbito, el principio implica que la recopilación y el procesamiento de datos deben llevarse a cabo de acuerdo con los parámetros y criterios establecidos por los humanos.(Guío Español et al., 2021).
- **Ética de los algoritmos:** Los algoritmos deben facilitar la toma de decisiones humanas. En principio, deben servir de guía para la toma de decisiones y no pueden actuar de forma automatizada y conforme a modelos sugeridos.(Guío Español et al., 2021).

- **Ética de las prácticas:** Los sistemas de IA no deben interactuar con la ciudadanía sin el control de un humano. Los sistemas de respuesta y conversación automatizada deben tener mecanismos para que los humanos puedan intervenir y participar en cualquier momento.(Guío Español et al., 2021).
- 

**Seguridad:** Este principio establece que los sistemas de inteligencia artificial (IA) deben proteger la integridad física y mental de los seres humanos con los que interactúan. La seguridad y la confidencialidad de los datos personales, especialmente los sensibles, son esenciales para evitar daños a la seguridad física y mental de los individuos.(Guío Español et al., 2021).

- **Ética de los datos:** En este contexto, se deben implementar mecanismos que garanticen la confidencialidad e integridad de la información utilizada por los sistemas de IA. Esta información no debe ser alterada en ningún momento. Se deben generar mecanismos para prevenir tales alteraciones de la información (aprendizaje adversarial de la máquina).(Guío Español et al., 2021).
  - **Ética de los algoritmos:** El diseño e implementación de algoritmos deben seguir un sistema de gestión de riesgos que permita identificar las posibles afectaciones que ciertos resultados pueden generar y la posibilidad de evitarlos. Un algoritmo nunca debe conducir a un resultado que ponga en riesgo la integridad de un ser humano. Las decisiones más trascendentales solo deben ser tomadas por humanos, y los algoritmos deben servir como modelos de guía en estas decisiones.(Guío Español et al., 2021).
  - **Ética de las prácticas:** Se deben evitar prácticas que pongan en riesgo los sistemas de IA. Los códigos de conducta deben establecer parámetros para prevenir actividades que pongan en peligro la integridad y la seguridad física de las personas.(Guío Español et al., 2021).
-

**Responsabilidad:** Este principio establece que existe un deber de responder por los resultados producidos por un sistema de inteligencia artificial y las afectaciones que pueda causar. Todos los actores involucrados en la cadena algorítmica tienen un papel y la responsabilidad no puede limitarse a un participante específico. En principio, se asume la solidaridad en la responsabilidad de los diseñadores, desarrolladores y personas que implementan esta tecnología por los daños causados a un individuo, a menos que se pueda demostrar que la responsabilidad recae en un solo actor.(Guío Español et al., 2021).

- **Ética de los datos:** Todas las entidades que recolectan y procesan datos para el diseño, desarrollo e implementación de sistemas de inteligencia artificial deben ser consideradas responsables de esta información. Deben responder por su integridad y las finalidades del procesamiento. En ningún caso, la responsabilidad debe recaer solo en uno de estos actores.(Guío Español et al., 2021).
- **Ética de los algoritmos:** Las personas que diseñan un algoritmo son responsables de los resultados que genera y los criterios utilizados para llegar a ciertas respuestas. Sin embargo, su responsabilidad no se extiende hasta la fase de implementación, ya que en esta etapa los resultados y su impacto son responsabilidad de la persona o entidad que usa estos sistemas y toma decisiones a partir de ellos.(Guío Español et al., 2021).
- **Ética de las prácticas:** Los actores involucrados en el desarrollo de la tecnología de IA deben establecer responsabilidades claras en la cadena de diseño, producción e implementación. Dentro de los equipos de trabajo, debe haber una distribución clara de funciones y responsabilidades en su desarrollo y cumplimiento. Se deben evitar prácticas y acuerdos que limiten la responsabilidad en contraposición a lo establecido en este principio.(Guío Español et al., 2021).

**No Discriminación:** Este principio sostiene que los sistemas de inteligencia artificial no pueden producir resultados o respuestas que amenacen el bienestar de un grupo específico o que limiten los derechos de poblaciones históricamente marginadas. Los sistemas de IA deben ser neutrales y no limitarse a un grupo específico por razón de sexo, raza, religión, discapacidad, edad u orientación sexual.(Guío Español et al., 2021).

- **Ética de los datos:** Los datos utilizados deben ser analizados de manera que se minimice al máximo la posibilidad de utilizar información que contenga prejuicios o sesgos, ya sea en su contenido, clasificación o uso. Se deben favorecer mecanismos que permitan realizar un análisis previo de un conjunto de datos y detectar posibles problemas.(Guío Español et al., 2021).
- **Ética de los algoritmos:** Los algoritmos deben ser capaces de responder a las necesidades e intereses de diferentes grupos poblacionales. El rendimiento de un algoritmo no puede estar limitado a un grupo poblacional específico. Se debe realizar un seguimiento constante de los falsos positivos y negativos que un sistema pueda generar, analizando cómo los criterios de sexo, raza, religión, discapacidad, edad u orientación sexual pueden afectar estos resultados.(Guío Español et al., 2021).
- **Ética de las prácticas:** En el diseño de los sistemas de IA debe participar un grupo diverso de la población. Se deben generar análisis de impacto que permitan detectar tempranamente cualquier tipo de discriminación y corregirla de manera oportuna. Debe existir un análisis constante de este impacto y considerar mecanismos para retirar inmediatamente sistemas que tengan efectos discriminatorios.(Guío Español et al., 2021).

---

**Inclusión:** Este principio aboga por la participación activa de poblaciones históricamente marginadas y diversas en el diseño, desarrollo, implementación y evaluación de los sistemas de inteligencia artificial utilizados en Colombia. El Estado debe utilizar sistemas de inteligencia artificial que hayan cumplido con criterios de inclusión y respondan a las necesidades propias y específicas de estos grupos.(Guío Español et al., 2021).

- **Ética de los datos:** La inclusión implica el uso de datos que sean representativos y procedentes de diferentes grupos sociales, ya sea para el diseño, entrenamiento o funcionamiento de estos sistemas. Para este fin, se debe aumentar la disponibilidad de conjuntos de datos de los grupos históricamente menos representados.(Guío Español et al., 2021).
  - **Ética de los algoritmos:** Las variables que han sido incluidas dentro del algoritmo deben reconocer los efectos que pueden tener en contextos particulares y la posibilidad de que se privilegie a un grupo específico, evitando este tipo de diseño.(Guío Español et al., 2021).
  - **Ética de las prácticas:** Los grupos encargados del diseño, desarrollo e implementación deben tener en cuenta distintos sectores de la sociedad. Se deben establecer comités de evaluación para evitar prácticas discriminatorias contra grupos como las mujeres, afrodescendientes, indígenas o miembros de la comunidad LGBTI+, entre otros. Colombia debe liderar esfuerzos para evitar el diseño de sistemas de inteligencia artificial que hagan que las mujeres sean vistas como asistentes personalizadas y como seres al servicio de los consumidores. Es deseable desarrollar sistemas de inteligencia artificial que adopten un enfoque neutral desde el punto de vista de género y que no discriminen en base al género de los individuos.(Guío Español et al., 2021).
- 

**Prevalencia de los Derechos de Niños, Niñas y Adolescentes:** Este principio subraya la importancia de considerar a los niños, niñas y adolescentes (NNA) como un grupo especial con características y derechos únicos que necesitan un análisis más profundo en cuanto al impacto de la inteligencia artificial (IA). La implementación de estrategias y sistemas de IA debe reflejar las necesidades y potencial de este grupo.(Guío Español et al., 2021).

- **Ética de los datos:** Los datos de NNA solo pueden ser utilizados en actividades que se relacionen con su interés superior y que tengan un impacto exclusivamente positivo en ellos.(Guío Español et al., 2021).

- **Ética de los algoritmos:** El diseño y desarrollo de los algoritmos debe ser comprensible para los NNA, especialmente cuando estos puedan impactar su desarrollo y bienestar. Se debe evitar el diseño de cualquier algoritmo que pueda perjudicar a los menores de edad, especialmente en prácticas como la intimidación y la discriminación (bullying).(Guío Español et al., 2021).
  - **Ética de las prácticas:** Los NNA deben ser considerados en el desarrollo de estos sistemas cuando sean propios de sus actividades, estableciendo mecanismos específicos de participación en una lógica de cocreación que también les permita evaluar el impacto que estos sistemas tienen en esta población. Se deben generar programas de capacitación y educación que les permita a los NNA conocer y entender las características de esta tecnología y sus implicaciones, destacando la formación ética.(Guío Español et al., 2021).
- 

**Beneficio Social:** Este principio sugiere que la inteligencia artificial (IA) debe usarse de manera que genere un beneficio social claro y determinable, ya sea reduciendo costos, aumentando la productividad, mejorando el bienestar público, facilitando la prestación de servicios públicos, protegiendo el medio ambiente, proporcionando entretenimiento, entre otros. Los sistemas de IA que persiguen otros fines deben desalentarse, especialmente en el sector público.(Guío Español et al., 2021).

- **Ética de los datos:** Debe priorizarse el acceso fácil a los datos y la infraestructura de datos públicos para el desarrollo de sistemas de IA que muestren un beneficio social claro, especialmente en el diseño de políticas públicas y la prestación de servicios públicos.(Guío Español et al., 2021).
- **Ética de los algoritmos:** Los modelos y algoritmos utilizados deben tener como fin último un resultado ligado a un fin socialmente reconocido. Por lo tanto, se debe demostrar cómo los resultados esperados se relacionan con dicho fin social.(Guío Español et al., 2021).
- **Ética de las prácticas:** Las personas que trabajan en el diseño, desarrollo e implementación de la IA en Colombia deben estar conscientes de las

principales dificultades sociales que enfrenta el país y cómo esta innovación puede ayudar a resolverlas. El Estado debe promover el uso de la IA en el proceso de transformación digital que busca reducir las brechas y disminuir la inequidad existente. Para esto, se deben establecer programas que promuevan retos públicos en el uso de IA (IAckaton) con el objetivo de solucionar problemas sociales específicos.(Guío Español et al., 2021).

---

Es fundamental identificar y comprender cada uno de los principios mencionados en el Marco Ético para la Inteligencia Artificial en Colombia (Guío Español et al., 2021) debido a su importancia y pertinencia en temas directamente relacionados con la sociedad, especialmente en aquellos contextos donde puede haber discriminación. Estos principios sirven como pilares esenciales para el desarrollo de la Inteligencia Artificial, y su adopción y seguimiento aseguran un enfoque equitativo, inclusivo y beneficioso para la sociedad en su conjunto. Al centrarse en estos principios, se pueden mitigar potenciales sesgos y desigualdades, y se promueve el uso ético y responsable de la Inteligencia Artificial.

#### 4.6.3. Herramientas para la implementación

#### ***TABLERO DE CONTROL 'DASHBOARD' DE SEGUIMIENTO AL MARCO ÉTICO PARA LA INTELIGENCIA ARTIFICIAL***

El Gobierno de Colombia ha establecido una herramienta innovadora para rastrear y monitorear la implementación de los principios establecidos en el "Marco Ético para la Inteligencia Artificial en Colombia" a través del 'Dashboard' de Seguimiento. Esta herramienta está diseñada para incluir proyectos de Transformación Digital e Inteligencia Artificial en todas las etapas de desarrollo, desde el diseño hasta la implementación y el despliegue. La singularidad de este modelo en Colombia es su aplicación a nivel nacional, lo cual subraya el compromiso del país con la ética de la IA en la región y en el mundo (Guío Español et al., 2021).

El 'Dashboard' de Seguimiento permite a las entidades públicas reportar cómo están implementando los principios éticos de la IA en sus proyectos. Los ciudadanos pueden

acceder a este tablero a través del sitio oficial del Gobierno de Colombia y conocer cómo cada proyecto se ajusta a los principios del Marco Ético, los Objetivos de Desarrollo Sostenible que cumple, y los retos que enfrenta, entre otros aspectos. Además, proporciona datos de contacto del equipo detrás del proyecto para que los ciudadanos puedan enviar comentarios o preguntas (Guío Español et al., 2021).

La herramienta requiere que las entidades informen y actualicen periódicamente una serie de información detallada sobre sus proyectos, incluyendo una descripción del proyecto y la problemática que intenta resolver, cómo la IA está siendo implementada en el proyecto, los Objetivos de Desarrollo Sostenible a los que contribuye el proyecto, y los mecanismos para la implementación de los principios éticos, entre otros. Esta transparencia es vital para acercar a la población a los proyectos de inteligencia artificial que se están desarrollando en el país (Guío Español et al., 2021).

En términos generales, el 'Dashboard' de Seguimiento sirve como un mecanismo invaluable para aumentar la transparencia en el desarrollo y la aplicación de los sistemas de IA, y para responder a la necesidad de los Estados de incrementar la transparencia en este ámbito. Esta herramienta es un paso importante hacia la mitigación de los desafíos que plantean los sistemas de IA para la sociedad y los derechos humanos (Guío Español et al., 2021).

## ***EVALUACIÓN DE ALGORITMOS***

La evaluación de algoritmos es una herramienta que ha sido explorada en los últimos años para evaluar la implementación de los principios éticos en el uso de datos y algoritmos en proyectos de inteligencia artificial (ACNUDH, 2021). Esta evaluación se consolida en informes periódicos que permiten un mapeo constante de los programas de IA del sector público, cómo se están implementando los principios éticos y los desafíos encontrados en este proceso (Stats New Zealand, 2018).

La autoridad de datos de Nueva Zelanda realizó una de las implementaciones más notables de este tipo de evaluación. A través de la revisión y análisis de casos de implementación de IA en entidades públicas, la autoridad busca garantizar que los ciudadanos

estén informados sobre el uso de algoritmos gubernamentales y los controles que existen para su manejo (Stats New Zealand, 2018).

Las conclusiones y recomendaciones de este informe proporcionan una oportunidad para que las agencias revisen y actualicen sus procesos para manejar los algoritmos. Esto puede ayudar a fomentar la innovación y las buenas prácticas en todo el sistema de datos, además de mejorar la transparencia y la responsabilidad (accountability) en el uso de algoritmos por parte del gobierno (Stats New Zealand, 2018).

Entre los hallazgos del informe, se encontró que, aunque algunas agencias del gobierno describen el uso de algoritmos con un estándar de buenas prácticas, no hay consistencia en todo el gobierno. También se identificaron oportunidades significativas para mejorar las descripciones de cómo los algoritmos informan o impactan la toma de decisiones (Stats New Zealand, 2018). La mayoría de las agencias gubernamentales no tienen procesos formales para revisar los algoritmos durante su desarrollo y operación, lo que sugiere un amplio margen para mejorar (Stats New Zealand, 2018).

A medida que la tecnología evoluciona, este continuará siendo un área en la que las agencias del gobierno deben equilibrar la importancia de la supervisión humana con las posibles eficiencias en la prestación de servicios (Stats New Zealand, 2018).

De acuerdo con estas recomendaciones, se propone que el gobierno colombiano considere la elaboración de este tipo de análisis y reportes. Esto permitiría tener un mapeo constante de los proyectos transformadores dentro del sector público en los que se utiliza esta tecnología, y la forma en que se están implementando los principios éticos dentro de la implementación y despliegue de esa tecnología (ACNUDH, 2021).

### **'LIMPIEZA' DE DATOS**

El proceso de 'limpieza' de datos es una medida crucial para limitar los sesgos y errores en los datos utilizados en el desarrollo de tecnologías basadas en inteligencia artificial. Este proceso comprende una serie de pasos diseñados para depurar, corregir y actualizar la información. Algunos de los pasos más destacados en este proceso incluyen (Gimenez, 2018):

- Monitoreo de errores: Es importante llevar un registro y observar las tendencias para identificar de dónde provienen la mayoría de los errores. Esto facilita la identificación y corrección de datos incorrectos o corruptos. Es especialmente importante si se integran otras soluciones con el software administrativo principal, para evitar que los errores obstaculicen el trabajo de otros departamentos.
- Estandarización de procesos: La estandarización del punto de entrada y su revisión son aspectos críticos. Al estandarizar el proceso de datos, se asegura un buen punto de entrada y se reduce el riesgo de duplicación.
- Validación de precisión: Una vez que se ha limpiado la base de datos existente, es importante validar la precisión de los datos. Para ello, se recomienda investigar e invertir en herramientas de datos que ayuden a limpiar los datos en tiempo real. Algunas herramientas incluso utilizan IA o machine learning para mejorar la precisión.
- Identificación de datos duplicados: Identificar los datos duplicados puede ahorrar tiempo al analizar los datos. Esto se puede lograr buscando y utilizando las herramientas de limpieza de datos, que pueden analizar datos en masa y automatizar el proceso.
- Análisis: Una vez que los datos han sido estandarizados, validados y revisados por duplicados, se deben utilizar terceros para agregar los datos. Las fuentes externas confiables pueden recolectar información de primera mano, luego limpiar y compilar los datos para proporcionar información más completa para inteligencia y análisis de negocios.
- Comunicación con el equipo: Es crucial comunicar el nuevo proceso estandarizado de limpieza al equipo. Una vez que los datos se han limpiado, es importante mantenerlos así. Esto ayudará a desarrollar y fortalecer la segmentación de los consumidores y a enviar información mejor dirigida a los consumidores y prospectos.

Esta medida es especialmente relevante en aquellos sectores que pueden ser susceptibles al uso de datos que pueden tener mayores prejuicios o que pueden estar "contaminados". También es una medida altamente recomendable cuando las entidades se enfrentan a bases de datos cuya calidad puede ser cuestionable (Gimenez, 2018).

## ***EXPLICACIÓN INTELIGENTE***

El principio de "explicación inteligente" presenta una solución potencial a uno de los mayores desafíos en el campo de la inteligencia artificial: la necesidad de explicaciones comprensibles para el funcionamiento de los sistemas de IA, que a menudo son incomprensibles incluso para los expertos (Doshi-Velez & Kortz, 2017). Este principio sugiere que las explicaciones deben proporcionarse en aquellos casos en los que los beneficios de la explicación superen sus costos.

Por lo tanto, es crucial considerar por qué y cuándo las explicaciones son lo suficientemente útiles para superar los costos. La obligación de que todos los sistemas de IA expliquen todas las decisiones podría resultar en sistemas menos eficientes, decisiones de diseño forzadas y un sesgo hacia resultados explicables pero insuficientes. Por ejemplo, el costo adicional de obligar a una tostadora a explicar por qué cree que el pan está listo podría prevenir a una empresa de implementar una función de tostadora inteligente debido a desafíos de ingeniería o preocupaciones legales (Doshi-Velez & Kortz, 2017).

Sin embargo, también puede ser aceptable asumir los costos económicos de un sistema de aprobación de créditos más explicable pero menos preciso por el beneficio social de poder verificar que no sea discriminatorio. Las normas sociales que determinan cuándo se necesitan explicaciones también deben aplicarse a los sistemas de IA (Doshi-Velez & Kortz, 2017).

La explicación inteligente evita que los sistemas de IA evadan el mismo nivel de escrutinio que se espera de los humanos, pero también previene la exigencia excesiva a los sistemas de IA hasta el punto de obstruir la innovación y el progreso. Este enfoque modesto plantea sus propios desafíos, y su implementación proporcionará una mejor idea de si y dónde los requisitos de explicación deben diferir para los sistemas de IA y los humanos (Doshi-Velez & Kortz, 2017).

Dado que tenemos pocos datos para determinar los costos reales de requerir que los sistemas de IA proporcionen explicaciones, el papel de la explicación en garantizar la responsabilidad también debe reevaluarse periódicamente para adaptarse al panorama siempre cambiante de la tecnología (Doshi-Velez & Kortz, 2017).

### ***EVALUACIÓN DE LA LEGITIMIDAD***

El modelo propuesto por la Web Foundation tiene como objetivo evaluar la legitimidad en la implementación de sistemas de inteligencia artificial, particularmente por parte de entidades públicas. La legitimidad en la implementación se logra cuando el proceso es explicable, responsabilidades pueden ser rastreadas, y se pueden identificar claramente quienes son los actores involucrados en las distintas operaciones de diseño y desarrollo de un sistema de inteligencia artificial. Además, los resultados deben ser no discriminatorios, justos, y debe ser posible determinar y minimizar el impacto de los falsos positivos y negativos (World Wide Web Foundation, 2018).

Basado en las declaraciones de la WWW Foundation en el año 2018, el marco ético se acoge a los 4 pasos explicados en la siguiente imagen:

Figura 10

*We suggest that public officials consider four key areas to assess the effectiveness and legitimacy of an AI system's implementation:*

**1. The process of dataset creation, e.g.:**

- Who determines what data to collect?
- Who is included within the data?

**2. The setup and design of AI tools, e.g.:**

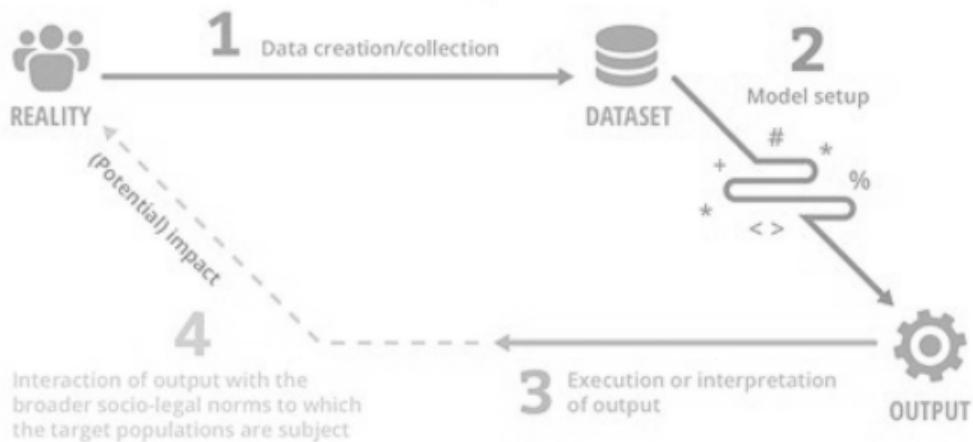
- What variables were included?
- Do they trigger risk of undue discrimination?
- Are outputs explainable? To whom? How?
- How do the outputs compare to human prediction or other non-AI systems?

**3. The administrative protocols that surround the tool's output, e.g.:**

- Does the tool inform human decisions, or execute policies directly?

**4. Interaction with broader social and legal norms target populations are subject to, e.g.:**

- Are there mechanisms of appeal for citizens who are impacted by decisions made by AI tools?
- What other safety-nets are available to those who are denied a service?
- How will the community treat a person who the AI classified in a certain way?



(World Wide Web Foundation, 2018)

### **DEFINICIÓN Y GESTIÓN DE RIESGOS ÉTICOS:**

En el contexto del despliegue de un modelo de Inteligencia Artificial (IA), las organizaciones deben adoptar un sistema coherente de gestión de riesgos y controles internos para abordar los riesgos asociados (IMDA & PDPC, 2020). Entre las medidas propuestas se incluyen:

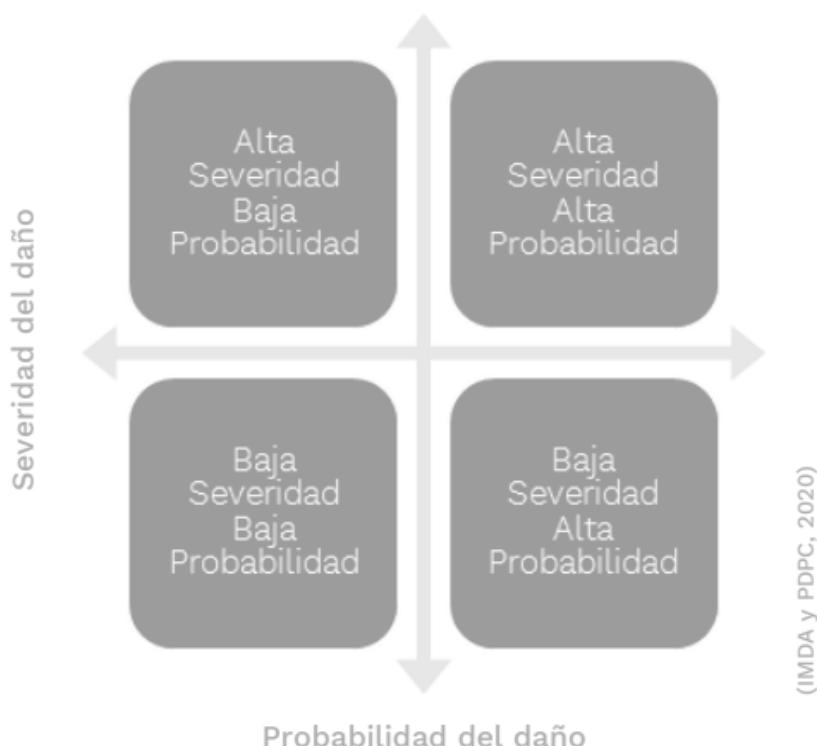
- La utilización de esfuerzos razonables para garantizar la idoneidad de las bases de datos utilizadas para el entrenamiento del modelo de IA. Esto implica la evaluación y gestión de los riesgos de inexactitud o sesgo y la revisión de las excepciones identificadas durante el entrenamiento del modelo (IMDA & PDPC, 2020).
- La instauración de sistemas de monitoreo y reporte, así como procesos para garantizar que el nivel administrativo competente esté informado del

rendimiento y otros asuntos relacionados con el sistema desplegado (IMDA & PDPC, 2020).

- Asegurar una apropiada transferencia de conocimientos en caso de cambios en el personal clave relacionado con las actividades de IA, lo que disminuirá el riesgo de que el movimiento del personal cree una brecha en la gobernanza interna (IMDA & PDPC, 2020).
- Revisar la estructura de gobernanza interna y las medidas cuando haya cambios significativos en la estructura organizacional o en el personal clave involucrado (IMDA & PDPC, 2020).
- Revisar periódicamente la estructura de gobernanza interna y las medidas para asegurar su relevancia y efectividad continuas (IMDA & PDPC, 2020).

Además, IMDA y PDPC (2020) proponen una matriz para clasificar la probabilidad y la severidad del daño a un individuo como resultado de una decisión tomada por una organización sobre ese individuo. La definición de daño, así como el cálculo de la probabilidad y la severidad, dependerá del contexto y variará de un sector a otro.

Figura 11



## **OTRAS HERRAMIENTAS**

Además de las herramientas ya discutidas, el marco ético sugiere varias estrategias adicionales. Estas incluyen la creación de códigos internos de conducta o ética, la realización de análisis de impacto en la privacidad, y la implementación de modelos de gobernanza ética en inteligencia artificial. Cabe destacar que cada una de estas estrategias ya se encuentra incorporada en el desarrollo de nuestro proyecto. Esto se debe al compromiso con el código de ética de nuestra universidad, la consideración de la naturaleza de los datos manejados, y la falta de un modelo de gobernanza regional existente. Esta última condición subraya la necesidad de un estándar que guíe un uso apropiado y responsable de la inteligencia artificial, de acuerdo con las recomendaciones del marco ético presentado.

### **4.7. Estado del Arte**

#### **4.7.1. Aplicaciones de IA para la predicción de crímenes en el sector académico**

El uso de la inteligencia artificial (IA) para la prevención y la predicción de crímenes ha ganado importancia en el sector académico en los últimos años. Por ejemplo, Plan Internacional y la empresa Inteligencia Artificial en Epidemiología Médica (AIME) han desarrollado "Eagle Eye", una plataforma de análisis de datos que emplea la IA y el aprendizaje automático para generar predicciones de feminicidios. El sistema, que fue alimentado con datos proporcionados por la Procuraduría General de la República de República Dominicana, genera predicciones de posibles casos de feminicidio hasta el año 2023. Aunque esta herramienta está aún en su fase de prueba, ha demostrado la capacidad de identificar los meses y los municipios con mayor probabilidad de registrar feminicidios.

Figura 12



#### 4.7.2. Uso de IA para la predicción de crímenes en el sector público

En el sector público, la Oficina de las Naciones Unidas contra la Droga y el Delito (UNODC) en México está utilizando la inteligencia artificial para prevenir la violencia contra las mujeres. El Centro de Excelencia UNODC-INEGI utiliza la IA para analizar los reportes de llamadas al 911 y detectar casos de violencia contra las mujeres. El sistema utiliza redes neuronales artificiales para el procesamiento del lenguaje natural, lo que le permite analizar la transcripción de las llamadas y detectar casos de violencia contra las mujeres que podrían haber sido clasificados bajo un tipo de violencia distinto.

Estos ejemplos demuestran cómo la inteligencia artificial puede ser utilizada para identificar patrones en grandes volúmenes de datos y predecir crímenes. Sin embargo, es importante tener en cuenta que estas herramientas dependen de la calidad y la precisión de los datos con los que se alimentan. La falta de información detallada sobre los perpetradores y la existencia de discrepancias en los datos pueden afectar la precisión de las predicciones.

Dado este escenario, es fundamental continuar mejorando la calidad de los datos y la precisión de los modelos de IA. Con un esfuerzo sostenido en esta dirección, podremos mejorar la efectividad de estas herramientas en la prevención y la predicción de crímenes.

**Matriz de Incidencia:**

Alternativa	Cumplimiento del marco ético	Consumo de Recursos	Formato de input/output	Uso de variables adicionales (No espacial/temporal)	Metodología que usan	Costo/Beneficio
Towards Intersectional Feminist and Participatory ML	Bajo	Medio	Medio	Bajo	Medio	Bajo
Implementación de un sistema web integrado con chatbot para combatir la violencia contra la mujer	Alto	Alto	Medio	Bajo	Medio	Medio
PredPol	Medio	Alto	Alto	Alto	Alto	Alto
Análisis de factibilidad para la implementación de un sistema informático de gestión como modelo de prevención de delitos basado en algoritmos de simulación predictivos en el Comando Rural de Policía de la ciudad de Milagro.	-	-	-	-	-	Alto
Este Proyecto	Alto	Alto	Alto	Medio	Alto	Alto

## 5. Desarrollo del proyecto

### 5.1. Metodología

El desarrollo de este proyecto se fundamentó en una combinación de las metodologías CRISP-DM, KDD y Ágil, que proporcionaron un marco sólido para estructurar y guiar el trabajo del equipo en todas las etapas.

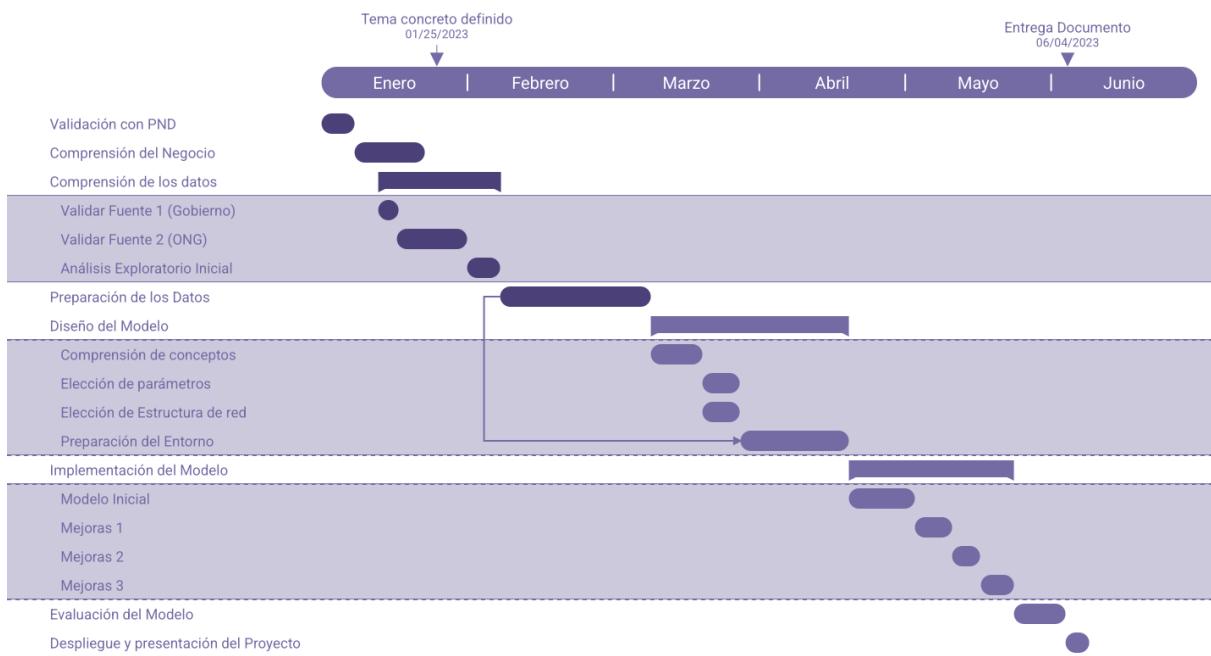
El CRISP-DM (Cross Industry Standard Process for Data Mining) es una metodología ampliamente utilizada en proyectos de minería de datos, diseñada para garantizar la comprensión completa del problema, el desarrollo adecuado del modelo y la evaluación exhaustiva del mismo. Esta metodología, estructurada en seis fases: Comprensión del negocio, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación y Despliegue, fue la base principal para el diseño y la implementación de nuestro proyecto.

La metodología KDD (Knowledge Discovery in Databases), por otro lado, se enfoca en el proceso de extracción de patrones significativos y conocimientos útiles a partir de grandes conjuntos de datos. Su énfasis en las fases de preprocesamiento de datos, selección, transformación, minería de datos y evaluación de los patrones descubiertos, complementó al CRISP-DM en el análisis y la preparación de los datos.

Finalmente, la metodología Ágil, originada en el desarrollo de software y caracterizada por su enfoque iterativo y flexible, se incorporó para facilitar la adaptabilidad ante los cambios y las necesidades emergentes del proyecto. Su énfasis en la retroalimentación continua, la mejora del producto y la adaptación al cambio nos permitió iterar y mejorar nuestro modelo a lo largo del proceso.

En resumen, el proyecto se desarrolló siguiendo las fases y los principios de estas metodologías, lo que implicó la comprensión inicial del problema y los datos disponibles, la preparación y el análisis de los datos, el diseño e implementación del modelo, y finalmente su evaluación y refinamiento. El ciclo de vida del proyecto se estructuró en sprints siguiendo el enfoque Ágil, cada uno de los cuales culminó con una revisión y una retrospectiva para evaluar el progreso y planificar los siguientes pasos.

En todas las etapas, se mantuvo una perspectiva ética, con especial atención a la minimización de sesgos, la transparencia y la protección de la privacidad de los datos. La privacidad y la confidencialidad de los datos se respetarán rigurosamente, y se tomaron precauciones para garantizar que el modelo resultante no perpetuara ni exacerbara los sesgos existentes.



## 5.2. Fases de desarrollo del proyecto

### 5.2.1. Obtención de datos

La participación en el evento Data Jam marcó el inicio de este proyecto de grado. El evento, enfocado en el análisis innovador de datos abiertos, tiene como meta generar nuevos conocimientos y aprendizajes que puedan traducirse en soluciones prácticas, ya sea para mejorar políticas públicas, informar decisiones, mejorar la veeduría, entre otros. En este contexto, se logró implementar un modelo inicial de Inteligencia Artificial a través de Google Data Studio para presentar los datos preliminares del proyecto. Sin embargo, no fue posible culminar el proyecto en ese instante debido a la falta de información adicional y más detallada. Además de la insuficiencia de datos, otro factor que influyó en la no concreción del proyecto durante el Data Jam fue la formulación inicial del modelo. Este, lamentablemente, no logró captar la atención ni generar el interés esperado durante el desarrollo del evento.

No obstante, el receso temporal en el proyecto permitió una reevaluación y reajuste de las estrategias. Así, la obtención de los datos se convirtió en la primera etapa crucial de esta nueva fase del proyecto. Durante este paso, se recopilaron datos espaciales y temporales relacionados con los feminicidios en Colombia, principalmente del Observatorio de Feminicidios y del Instituto de Medicina Legal y Ciencias Forenses. Para proteger la privacidad de las personas involucradas, no se recogieron datos personales sobre las víctimas. En cambio, se priorizó la recolección de información de carácter más general, como la ubicación y el tiempo de los incidentes.

#### 5.2.2. Análisis y preparación de datos para el modelado

Una vez obtenidos los datos, se realizó un análisis preliminar para comprender mejor su estructura, características y posibles inconsistencias. Este análisis incluyó la comparación de los datos de las diferentes fuentes (el Observatorio de Feminicidios y el Instituto de Medicina Legal y Ciencias Forenses) para identificar discrepancias y huecos en la información registrada. Se encontró que existen casos de feminicidio registrados en los datos del Instituto de Medicina Legal y Ciencias Forenses que no aparecen en los datos del Observatorio de Feminicidios. Adicionalmente, se observó que el Observatorio distingue entre feminicidio y transfeminicidio, una categorización que no es utilizada por el Instituto de Medicina Legal y Ciencias Forenses.

Posteriormente, se procedió a la preparación de los datos para su modelado. Esta etapa es crucial en cualquier proyecto de ciencia de datos, ya que la calidad y el formato de los datos pueden tener un gran impacto en la efectividad del modelo resultante. La preparación de los datos implicó tareas como la limpieza de los datos (por ejemplo, manejo de valores faltantes o erróneos), la transformación de los datos en un formato adecuado para el modelado (como la codificación de fechas y horas, la normalización de las coordenadas espaciales, etc.), y la división de los datos en conjuntos de entrenamiento y prueba.

Una vez obtenidos y analizados los datos, se procedió a la etapa de preparación de los datos. En esta fase se llevó a cabo la limpieza, transformación y validación de los datos para que estuvieran listos para el modelado. Este proceso implicó múltiples pasos y técnicas, alineados con la metodología CRISP-DM y KDD.

**Limpieza y Transformación:** La limpieza de datos implicó la eliminación de entradas no relevantes o incorrectas, la corrección de errores, y la gestión de datos faltantes. La transformación de los datos implicó cambiar su formato, estructura o valores para prepararlos para el análisis y el modelado.

**Validación y Corrección de Datos Espaciales:** Como parte del proceso de preparación, se realizó una validación de los datos espaciales utilizando un archivo shapefile obtenido de la página del Departamento Administrativo Nacional de Estadística (DANE), que contiene información geográfica oficial de Colombia a nivel municipal y departamental.

Se compararon los datos espaciales de nuestras fuentes con los del shapefile del DANE. Durante esta comparación, se encontraron discrepancias a nivel municipal y departamental en los datos proporcionados por el Observatorio de Feminicidios. Se realizó un proceso de corrección de estas discrepancias para asegurar la precisión y coherencia de los datos espaciales.

### 5.2.3. Fase de diseño

#### 5.2.3.1. Diseño del modelo ConvLSTM-cGAN

La fase de diseño del modelo se enfocó en la implementación de un modelo ConvLSTM-cGAN, aprovechando así la capacidad del ConvLSTM para manejar datos secuenciales y la habilidad de una cGAN para generar instancias de datos condicionales. A continuación, se describen las etapas claves y decisiones tomadas durante este proceso:

#### **Arquitectura del modelo cGAN:**

La arquitectura del modelo se compone de dos partes principales: el generador y el discriminador.

1. El generador toma un vector latente y una etiqueta como entradas, la etiqueta proporciona información adicional sobre los factores condicionales (Ubicación geográfica y tiempo) y genera una secuencia que representa la tasa de feminicidios en los municipios de Colombia. Adicionalmente el generador, utiliza una capa de incrustación para transformar las etiquetas de clase en vectores, que luego se combinan con el

vector latente. Este vector combinado se pasa a través de una serie de capas Conv1D, diseñadas para aprender patrones espaciales y temporales en los datos. Finalmente se aplica una función de activación sigmoide para garantizar que la salida del generador esté en el mismo rango que los datos reales.

2. En nuestra estructura cGAN, el discriminador recibe una secuencia y una etiqueta, y su objetivo es distinguir entre secuencias reales (proveniente del conjunto de datos) y secuencias falsas (generadas por el generador). Al igual que el generador, el Discriminador utiliza una capa de incrustación para transformar las etiquetas de clase en vectores. Estos vectores luego se redimensionan para que coincidan con la forma de las secuencias de entrada y se concatenan con ellas.

Este conjunto de datos combinado se pasa a través de varias capas convolucionales unidimensionales (Conv1D). Estas capas están diseñadas para identificar patrones temporales en las secuencias de entrada que pueden indicar si son reales o generadas. Finalmente, se aplica una función de activación sigmoide a la salida del Discriminador, esto asegura que la salida del Discriminador pueda interpretarse como una probabilidad, que es esencial para el entrenamiento del modelo.

#### 5.2.3.2. Elección de parámetros y estructura de la red

La elección de parámetros y la estructura de la red son dos de las decisiones más cruciales en el diseño de cualquier modelo de aprendizaje profundo. Estas decisiones a menudo requieren un conocimiento profundo del problema, así como experimentación.

A continuación mostraremos qué parámetros elegimos para el entrenamiento del modelo cGAN:

**Longitud de las secuencias de tiempo (seq\_length):** Seleccionamos una longitud de secuencia de 42. Este valor representa la cantidad de datos temporales consecutivos que se alimentarán al modelo para realizar cada predicción. La elección de este valor es esencial, ya que una longitud de secuencia demasiado corta puede no proporcionar suficiente contexto para las predicciones, mientras que una demasiado larga puede hacer que el modelo sea más difícil de entrenar y pueda provocar un sobreajuste.

**Número de municipios (num\_municipios):** Este parámetro se refiere al número de diferentes municipios presentes en nuestro conjunto de datos. Como queremos realizar predicciones para cada municipio, este número determina la longitud de los vectores en nuestras secuencias, en total son 1121 municipios.

**Dimensión latente (latent\_dim):** Este valor representa el tamaño del espacio de representación latente, que es el espacio desde el cual el generador seleccionará puntos para generar nuevas secuencias. Elegimos una dimensión latente de 20 para equilibrar la capacidad del modelo para generar diversidad y su capacidad para aprender representaciones útiles.

**Tamaño del lote (batch\_size):** Decidimos establecer un tamaño de lote de 128, equilibrando la necesidad de proporcionar suficientes ejemplos para que el modelo aprenda de cada actualización de los pesos y la limitación de la cantidad de memoria que se puede utilizar a la vez.

**Número de épocas (num\_epochs):** Seleccionamos un total de 100 épocas para el entrenamiento de nuestro modelo. Esto representa el número de veces que todo el conjunto de datos pasa por la red durante el entrenamiento.

**Número de categorías (n\_cats):** Finalmente, el número de categorías es igual al número de municipios, ya que queremos que nuestro modelo sea capaz de generar datos para cada uno de ellos.

```
# Parámetros Iniciales con el data Loader
seq_length = 42 # Longitud de las secuencias de tiempo
num_municipios = len(municipios) # Longitud de los vectores en tus secuencias
latent_dim = 20
batch_size = 128
num_epochs = 100
n_cats = len(municipios)

# Crear modelos y optimizadores
generator = build_cgan_generator(latent_dim, n_cats)
discriminator = build_cgan_discriminator(seq_length, num_municipios, n_cats)

generator.summary()
discriminator.summary()

generator_optimizer = adam_v2.Adam(learning_rate=0.0002, beta_1=0.5)
discriminator_optimizer = adam_v2.Adam(learning_rate=0.0002, beta_1=0.5)

# Crear dataloader
#data_loader = dataloader(input_sequences, batch_size)
data_gen = data_loader(input_sequences, municipios, batch_size, municipio_a_indice)
```

Adicional los comandos Adams que se ven en la Figura X son utilizados para configurar los optimizadores para el generador y discriminador de la cGAN. Adam es eficaz debido a su manejo adaptativo de las tasas de aprendizaje individuales para cada parámetro en el modelo. Esto significa que ajusta la tasa de aprendizaje para cada parámetro de forma individual, lo que puede ser muy beneficioso para la convergencia del modelo. Dentro del marco de trabajo del algoritmo de optimización Adam, existen dos parámetros esenciales que requieren una cuidadosa atención y configuración. Estos parámetros son: la tasa de aprendizaje (learning rate) y el coeficiente de decaimiento del primer momento (beta 1).

- Learning rate=0.0002: Esta es la tasa de aprendizaje y juega un papel crucial en la determinación del tamaño del paso en cada iteración del proceso de optimización del gradiente descendente. Es un hilo delicado a equilibrar, dado que una tasa de aprendizaje más baja implica que el modelo requerirá más tiempo para aprender; sin embargo, ofrece una ventaja: permite una convergencia más estable y reduce el riesgo de sobrepasar el mínimo global durante el proceso de optimización. Por otro lado, una tasa de aprendizaje más alta podría acelerar el proceso de aprendizaje, pero a la vez puede conducir a una convergencia inestable o incluso divergente.
- beta\_1=0.5: Este es un parámetro esencial que controla el coeficiente de decaimiento para la estimación del primer momento en el algoritmo Adam. Aunque el valor por defecto de beta 1 en Adam es 0.9, es común utilizar un valor de 0.5 cuando se entrena una GAN. Este parámetro tiene la responsabilidad de controlar la envergadura de los ajustes a los pesos durante cada actualización del modelo. En esencia, beta 1 determina la tasa de decaimiento para el promedio móvil del primer momento, que corresponde a la magnitud de la actualización del gradiente. Un valor más bajo para beta 1 se traducirá en dar más peso a los gradientes más recientes, permitiendo un ajuste más reactivo a las nuevas tendencias de los datos.

Un aspecto importante de nuestra implementación es el uso de optimizadores Adam separados para el generador y el discriminador. Esto ofrece la posibilidad de permitir que cada componente de la GAN se actualice a su propio ritmo durante el proceso de entrenamiento. Este aspecto es crucial, dado que el generador y el discriminador no suelen

aprender al mismo ritmo en las GANs, y ajustar sus optimizadores de forma independiente nos proporciona mayor control para mantener el equilibrio en el sistema de aprendizaje competitivo que constituye una GAN.

#### 5.2.4. Fase de implementación

Esta fase comprende la traducción del diseño del modelo en código utilizando la biblioteca TensorFlow y el entrenamiento del modelo. Aquí se detalla cada paso:

##### 5.2.4.1. Codificación y entrenamiento del modelo en TensorFlow

La implementación del modelo ConvLSTM-cGAN en código se llevó a cabo en Python, utilizando la biblioteca de aprendizaje profundo TensorFlow, debido a su flexibilidad y la gran cantidad de características que ofrece para la implementación de modelos de redes neuronales complejas. Las funciones **build\_cgan\_generator** y **build\_cgan\_discriminator** se utilizaron para implementar las arquitecturas del generador y del discriminador, respectivamente. Ambos componentes se construyeron utilizando la API funcional de Keras en TensorFlow, que permite una mayor flexibilidad en la conexión de capas no secuenciales, algo esencial para las arquitecturas GAN. Adicional se muestra el código a continuación:

```
# Generador
def build_cgan_generator(latent_dim, n_cats):
    # Label Inputs
    in_label = layers.Input(shape=(1,))
    lbls = layers.Embedding(n_cats, 50)(in_label)
    lbls = layers.Dense((seq_length) * num_municipios)(lbls)
    lbls = layers.Reshape([seq_length, num_municipios])(lbls) # Ajustamos las dimensiones

    # Generator Inputs (latent vector)
    in_latent = layers.Input(shape=(latent_dim,))
    gen = layers.Dense((seq_length) * num_municipios)(in_latent)
    gen = layers.Reshape([seq_length, num_municipios])(gen) # Ajustamos las dimensiones

    # Combine both inputs
    concat = layers.concatenate([gen, lbls])

    # Hidden Layer 1
    g = layers.Conv1D(filters=128, kernel_size=3, padding='same')(concat)
    g = layers.LeakyReLU(alpha=0.2)(g)

    # Hidden Layer 2
    g = layers.Conv1D(filters=128, kernel_size=3, padding='same')(g)
    g = layers.LeakyReLU(alpha=0.2)(g)

    # Output Layer
    output_layer = layers.Dense(num_municipios, activation='sigmoid')(g)

    # Define model
    model = Model([in_latent, in_label], output_layer)
    return model
```

```

# Discriminador
def build_cgan_discriminator(seq_length, num_municipios, n_cats):
    # Image input
    in_img = layers.Input(shape=(seq_length, num_municipios))

    # Label input
    in_label = layers.Input(shape=(1,))
    lbls = layers.Embedding(n_cats, 50)(in_label)
    lbls = layers.Dense((seq_length) * num_municipios)(lbls)
    lbls = layers.Reshape([(seq_length, num_municipios)])(lbls)

    # Concatenate image and label inputs
    concat = layers.concatenate([in_img, lbls])

    # Hidden Layer 1
    d = layers.Conv1D(filters=128, kernel_size=3, padding='same')(concat)
    d = layers.LeakyReLU(alpha=0.2)(d)

    # Hidden Layer 2
    d = layers.Conv1D(filters=128, kernel_size=3, padding='same')(d)
    d = layers.LeakyReLU(alpha=0.2)(d)
    d = layers.Flatten()(d)

    # Output Layer
    output_layer = layers.Dense(1, activation='sigmoid')(d)

    # Define model
    model = Model([in_img, in_label], output_layer)
    return model

```

```

#Modelo optimizado PRUEBA Urcuqui
class CGAN(keras.Model):
    def __init__(self, discriminator, generator, latent_dim):
        super(CGAN, self).__init__()
        self.discriminator = discriminator
        self.generator = generator
        self.latent_dim = latent_dim

    def compile(self, d_optimizer, g_optimizer, loss_fn):
        super(CGAN, self).compile()
        self.d_optimizer = d_optimizer
        self.g_optimizer = g_optimizer
        self.loss_fn = loss_fn

    def train_step(self, real_sequences):
        if isinstance(real_sequences, tuple):
            real_sequences = real_sequences[0]

        # Sample random points in the latent space
        batch_size = tf.shape(real_sequences)[0]
        random_latent_vectors = tf.random.normal(shape=(batch_size, self.latent_dim))

        # Decode them to fake sequences
        generated_sequences = self.generator.fit(random_latent_vectors, steps_per_epoch=10)

        # Combine them with real sequences
        combined_sequences = tf.concat([generated_sequences, real_sequences], axis=0)

        # Assemble labels discriminating real from fake sequences
        labels = tf.concat([
            tf.ones((batch_size, 1)) * 0.9, tf.zeros((batch_size, 1)) * 0.1], axis=0)
        ) # Aplicamos el suavizado de etiquetas

        # Train the discriminator
        with tf.GradientTape() as tape:
            predictions = self.discriminator.fit(combined_sequences, steps_per_epoch=10)
            d_loss = self.loss_fn(labels, predictions)
            grads = tape.gradient(d_loss, self.discriminator.trainable_weights)
            self.d_optimizer.apply_gradients(zip(grads, self.discriminator.trainable_weights))

        # Sample random points in the latent space
        random_latent_vectors = tf.random.normal(shape=(batch_size, self.latent_dim))

        # Assemble labels that say "all real sequences"
        misleading_labels = tf.ones((batch_size, 1)) * 0.9

        # Train the generator (note that we should *not* update the weights
        # of the discriminator)!
        with tf.GradientTape() as tape:
            predictions = self.discriminator(self.generator(random_latent_vectors))
            g_loss = self.loss_fn(misleading_labels, predictions)
            grads = tape.gradient(g_loss, self.generator.trainable_weights)
            self.g_optimizer.apply_gradients(zip(grads, self.generator.trainable_weights))

        return {"d_loss": d_loss, "g_loss": g_loss}

```

Para afrontar el desafío de la representación de los datos categóricos (municipios), se utilizó una capa de incrustación que mapea cada municipio a un vector denso de tamaño fijo, que luego se dimensiona a la forma requerida para la entrada de la red. Las capas de convolución 1D se utilizan para aprender patrones locales a lo largo del eje temporal de las secuencias. Finalmente, la clase CGAN une el generador y el discriminador y define el paso de entrenamiento personalizado necesario para un GAN, donde primero se entrena el

discriminador y luego el generador en tandem, usando los gradientes y las predicciones del discriminador.

#### 5.2.4.2. Iteraciones y mejoras del modelo

El proceso iterativo de entrenamiento es esencial para mejorar la eficiencia y precisión de los modelos generativos. En cada iteración o época, el generador crea secuencias ficticias a partir de vectores de ruido aleatorios y etiquetas categóricas. Estos vectores de ruido siguen una distribución aleatoria normalmente y actúan como la entrada para el generador. Mientras tanto, el discriminador tiene la tarea de distinguir entre las secuencias reales y las generadas por el generador. El generador y el discriminador se entrena alternativamente, y el rendimiento de cada uno de ellos influye en el aprendizaje del otro.

A medida que se desarrolla el entrenamiento, el generador se vuelve cada vez más eficaz en la producción de secuencias que engañan al discriminador. Del mismo modo, el discriminador mejora continuamente su capacidad para distinguir entre secuencias verdaderas y falsas. El proceso iterativo se realiza utilizando la función de pérdida de entropía cruzada binaria, que mide la eficacia con la que el discriminador puede clasificar secuencias y la eficacia con la que el generador puede engañar al discriminador, en este caso pasando información aleatoria o falsa sobre los municipios de Colombia donde se llevó a cabo el suceso del crimen de Feminicidio en Colombia.

### **Fase de Iteraciones y Mejoras del Modelo**

En esta fase, se realizan varias iteraciones de entrenamiento y se refinan y mejoran los modelos basándose en el rendimiento observado.

#### **1. Ajuste de Hiper Parámetros:**

El ajuste de los hiper parámetros del modelo es una parte crucial de esta fase. Los hiper parámetros como la tasa de aprendizaje, la dimensión del espacio latente y el tamaño del lote se ajustan y optimizan para mejorar el rendimiento del modelo. Este proceso es típicamente iterativo y requiere experimentación y validación cruzada.

## **2. Mejoras en la Arquitectura del Modelo:**

En función del rendimiento del modelo y de la interpretación de las métricas de pérdida, se pueden hacer ajustes en la arquitectura del modelo. Por ejemplo, se pueden agregar o eliminar capas, o se pueden cambiar los tipos de capas utilizadas (por ejemplo, cambiar entre Conv1D y LSTM dependiendo de la estructura de los datos). Además, se pueden probar diferentes técnicas de regularización como el dropout o la normalización por lotes para mejorar la capacidad del modelo de generalizar a partir de los datos.

## **3. Refinamiento del Proceso de Entrenamiento:**

Dependiendo de las dificultades encontradas durante el entrenamiento, se pueden introducir técnicas para mejorar la estabilidad y la eficiencia del entrenamiento. Algunas de estas técnicas pueden incluir:

- **Suavizado de etiquetas:** Esta es una técnica común para prevenir que el Discriminador se vuelva demasiado poderoso comparado al Generador. Consiste en añadir ruido a las etiquetas verdaderas y falsas usadas en el entrenamiento del Discriminador.
- **Clipping de los pesos:** Esta es otra técnica para mantener la estabilidad durante el entrenamiento de las GANs, que limita los valores de los pesos del discriminador dentro de un rango determinado.
- **Entrenamiento escalonado:** En lugar de entrenar al Generador y al Discriminador al mismo tiempo, podríamos entrenar al Discriminador unas cuantas épocas y luego al Generador. Esto puede ayudar a que el Discriminador no sea demasiado potente y que el Generador tenga tiempo para aprender a generar secuencias realistas.
- **Tasa de aprendizaje decreciente:** También se podría experimentar con disminuir la tasa de aprendizaje a medida que se avanza en las épocas de entrenamiento, lo que podría llevar a un entrenamiento más estable y preciso a largo plazo.

Estas mejoras e iteraciones se llevan a cabo hasta que se alcanza un punto de rendimiento aceptable o se cumple el límite de tiempo o recursos para el proyecto. Después de cada iteración y mejora, el modelo se vuelve a entrenar y se evalúa su rendimiento. Es importante destacar que cada uno de estos pasos requiere no sólo un conocimiento técnico profundo de cómo funcionan las redes neuronales y las GANs, sino también una gran cantidad de experimentación y ajuste fino. Aunque existen prácticas recomendadas y guías generales para el diseño y entrenamiento de estas redes, cada conjunto de datos y problemas tiene sus propios desafíos y peculiaridades únicas que deben ser atendidos. Por lo tanto, a pesar de que el proceso pueda parecer bastante técnico y mecánico, también requiere un alto nivel de habilidad creativa y de resolución de problemas.

### 5.2.5. Fase de validación

#### 5.2.5.1. Evaluación del modelo y ajustes necesarios

Es importante entender por qué el modelo de cGAN (Generative Adversarial Networks condicionales) no ha resultado ser la elección óptima para este proyecto. Las redes GAN son capaces de generar datos sintéticos muy realistas, y las cGAN añaden la capacidad de dirigir esa generación hacia ciertas características de interés. Sin embargo, el éxito de estos modelos depende en gran medida de la cantidad y calidad de los datos disponibles. En este caso, la dependencia de otras entidades para obtener la información ha representado un desafío significativo, y el modelo no ha alcanzado los resultados esperados.

Además, una limitación considerable en la implementación de este modelo ha sido la falta de acceso a equipos de cómputo potentes. El entrenamiento de modelos como las cGAN, que son inherentemente intensivos en cuanto a recursos de computación, puede requerir una capacidad de procesamiento y memoria significativa para ejecutarse de manera eficiente y efectiva. Esto ha dificultado la implementación y entrenamiento del modelo cGAN de manera óptima, teniendo en cuenta los recursos disponibles.

Por lo tanto, tanto la dependencia de entidades externas para la obtención de datos, como la limitación en los recursos computacionales, han representado obstáculos significativos en el desarrollo del proyecto. Esto ha llevado a la necesidad de replantear y ajustar el enfoque del proyecto para adecuarlo a estas restricciones.

### 5.2.6. Primera fase de ajustes al diseño, implementación y validación al modelo planteado

Tras implementar el modelo cGAN para analizar los feminicidios en Colombia, decidimos adoptar una estrategia innovadora que nos permitiera obtener resultados tanto en la etapa de preprocesamiento como en la de posprocesamiento de los datos. Esto implicaba un análisis exhaustivo de la escala de los datos, la evaluación de los intervalos de confianza y la automatización de la validación del modelo. En este escenario, decidimos utilizar las redes neuronales de tipo LSTM para realizar un análisis detallado del municipio de Bello, en el departamento de Antioquia, Colombia.

Nuestro primer paso fue la exploración de los datos. Identificamos la posibilidad de evaluar un municipio específico en relación con su departamento, para entender la frecuencia con la que ocurren los feminicidios en dicho lugar. De este modo, se seleccionó un intervalo temporal desde 2018 hasta 2023 para llevar a cabo nuestro análisis, contemplando la viabilidad de hacerlo con todos los departamentos y municipios.

La organización de la información se llevó a cabo de tal manera que todas las fechas dentro del rango de estudio quedaron representadas. Para aquellos días en los que no se registraron feminicidios, se asignó un valor de cero, garantizando así la continuidad en los datos.

Para dar inicio al proceso, se configuró un 'Forecaster' utilizando como referencia la fecha mínima y máxima del conjunto de datos correspondiente al departamento de Antioquia-Bello, desde el 2 de enero de 2018 hasta el 28 de enero de 2023. En este marco, se estableció el número de casos de feminicidios como variable dependiente y las fechas actuales como fechas de prensa del conjunto de datos. Con el objetivo de desarrollar un modelo que se ajuste a las especificidades de estos datos, se estableció una cadena de procesamiento que incluye etapas de preprocesamiento y modelización. El preprocesamiento se realiza mediante el uso de un 'TSColumnTransformer', que permite aplicar transformadores a columnas específicas de los datos de series temporales.

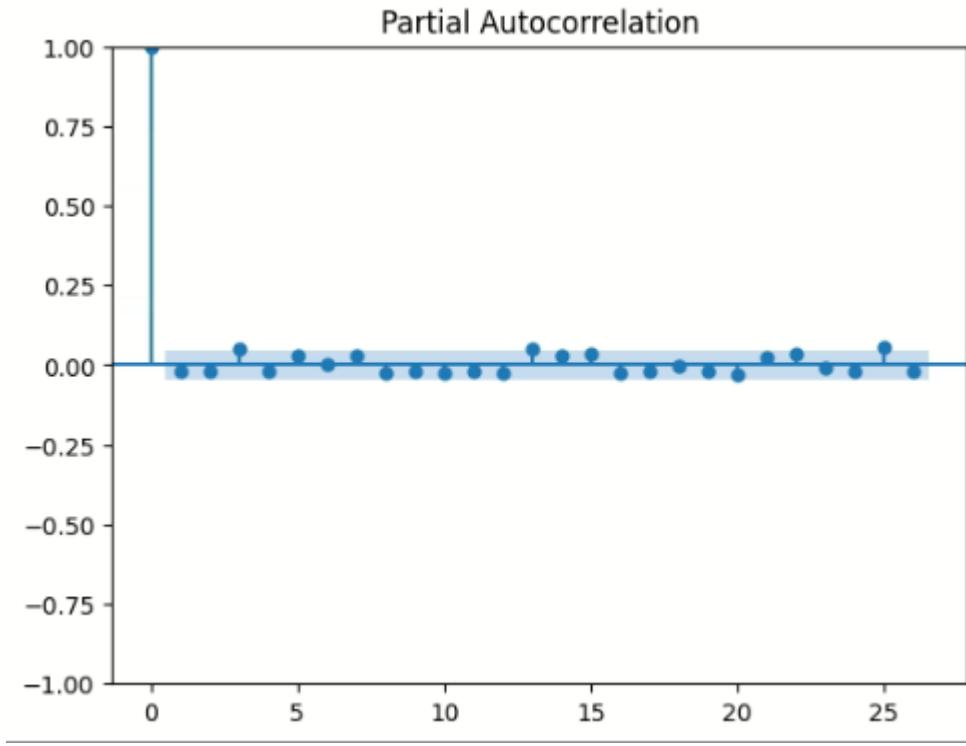
En este caso, se aplicaron dos transformadores: el 'SeasonalityTransformer', que se utiliza para eliminar la estacionalidad diaria de la variable objetivo, y el 'HolidayTransformer', que se utiliza para ajustar los datos teniendo en cuenta los días festivos en Colombia. Una vez pre-

procesados los datos, se implementa un modelo de Regresión Forestal Aleatoria, encapsulado en una envoltura de sklearn para permitir su integración en la cadena de procesamiento. Este modelo se seleccionó por su capacidad para manejar datos de alta dimensionalidad y su robustez ante posibles overfitting:

```
# Define model
model = Pipeline(
    steps=[
        (
            "preprocessing",
            TSColumnTransformer(
                transformers=[
                    ("seasonality", SeasonalityTransformer(freq="D"), ["target"]),
                    ("holidays", HolidayTransformer(country_code="CO"), ["target"])
                ]
            ),
        ),
        (
            "model",
            get_sklearn_wrapper(RandomForestRegressor, random_state=42),
        ),
    ]
)

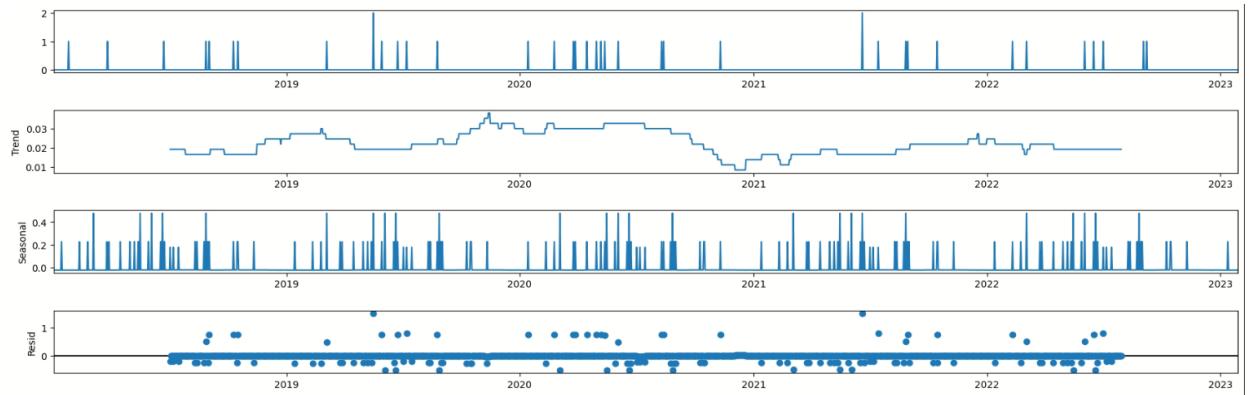
f = Forecaster(
    y = f_depar_mun['num_casos'],
    current_dates=f_depar_mun['fecha_en_prensa'],
    DateStartActuals = fecha_mas_antigua,
    DateEndActuals = ultima_fecha,
    Freq = "D",
    ForecastLength = 356,
    Xvars = [],
    Differenced = 0,
    TestLength = .4,
    ValidationLength = 4,
    ValidationMetric=['rmse', 'r2'],
    CILevel = 0.6,
    BootstrapSamples = 5000
)
```

Continuando con nuestro análisis, descompusimos esta serie temporal y visualizamos su gráfico de Función de Autocorrelación Parcial (PACF). Este análisis nos permitió medir el grado de correlación de la variable y (en nuestro caso, el número de casos de feminicidio en Bello, Antioquia) con sus valores anteriores y evaluar la existencia de una correlación estadísticamente significativa:



Nota. Mostrar la figura del PACF.

Además, desglosamos la serie temporal en sus componentes de tendencia, estacionalidad y residual. De este modo, pudimos analizar más a fondo los patrones y las variaciones inherentes en los datos:



Para verificar si la serie de tiempo era estacionaria o no, empleamos pruebas de estacionariedad, encontrando que nuestra serie sí lo era:

```

criticalpval = 0.05

print('*'*100)
print('Antioquia Augmented Dickey-Fuller results:')
stat, pval, _, _, _ = f.adf_test(full_res=True)
print('the test-stat value is: {:.2f}'.format(stat))
print('the p-value is {:.4f}'.format(pval))
print('the series is {}'.format('stationary' if pval < criticalpval else 'not stationary'))
print('*'*100)

-----
Antioquia Augmented Dickey-Fuller results:
the test-stat value is: -24.04
the p-value is 0.0000
the series is stationary
-----

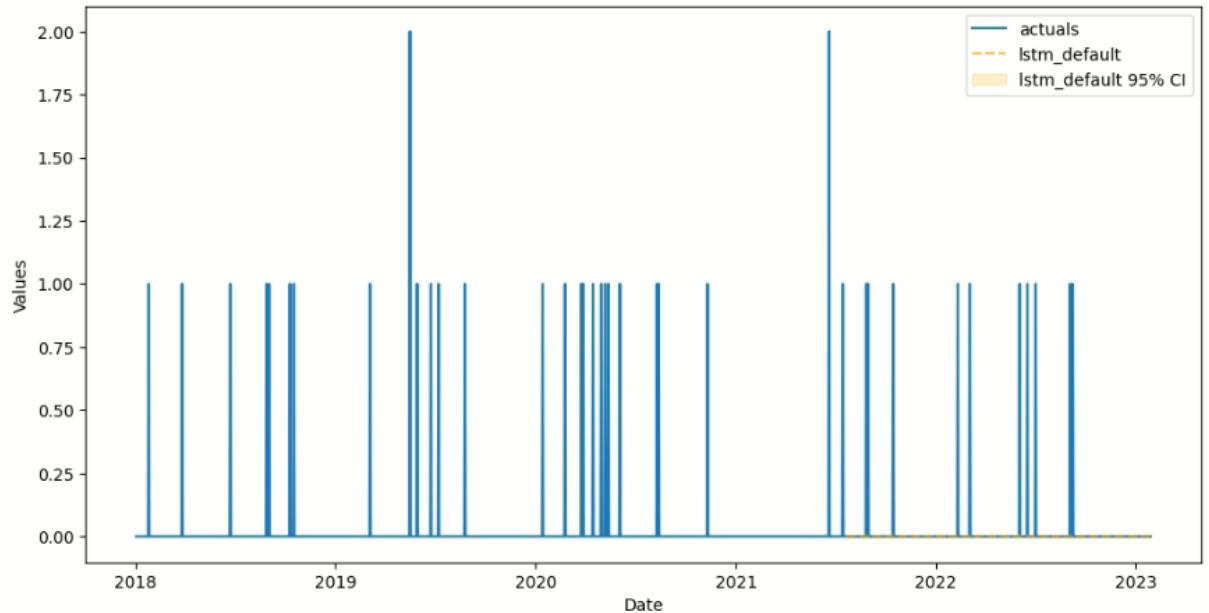
```

Definimos que el 30% del total de los datos serían utilizados para la validación del modelo. Además, establecimos 365 días como el horizonte de pronóstico y optamos por la red LSTM como nuestro modelo predictivo.

Al invocar una predicción LSTM, el modelo predeterminado se inicia con una única capa de entrada de tamaño 8. Emplea el optimizador Adam y una función de activación 'tanh'. El modelo se entrena por defecto con un único valor de la variable dependiente con retraso. Adicionalmente, este modelo adopta una tasa de aprendizaje de 0.001 y no incluye ninguna técnica de "dropout" para regularizar el modelo. Cabe destacar que todos los datos que ingresan al modelo son escalados a través de una técnica de escalado min-max, que es un método de normalización que transforma los valores a un rango entre 0 y 1. Sin embargo, al salir del modelo, los datos se presentan en su escala original, lo que significa que el escalado no es una transformación permanente sino solo una etapa temporal durante el procesamiento de datos.

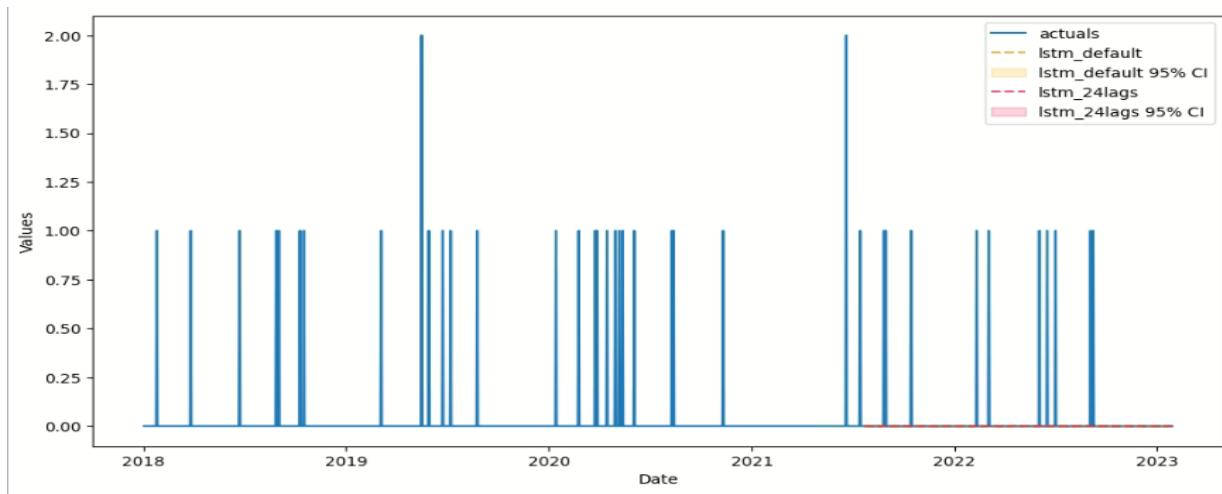
Es importante mencionar que cualquier argumento que se pueda pasar al método fit() en TensorFlow también es aceptado por el método manual forecast() de la librería Scale Cast. Esto proporciona una gran flexibilidad al permitirte personalizar el proceso de entrenamiento de acuerdo con las necesidades específicas de tu problema de series temporales.

Así, esta configuración proporciona un punto de partida sólido para modelar series de tiempo con redes LSTM, al tiempo que ofrece flexibilidad para ajustar los parámetros del modelo y mejorar la precisión de la predicción, el cual se muestra a continuación:

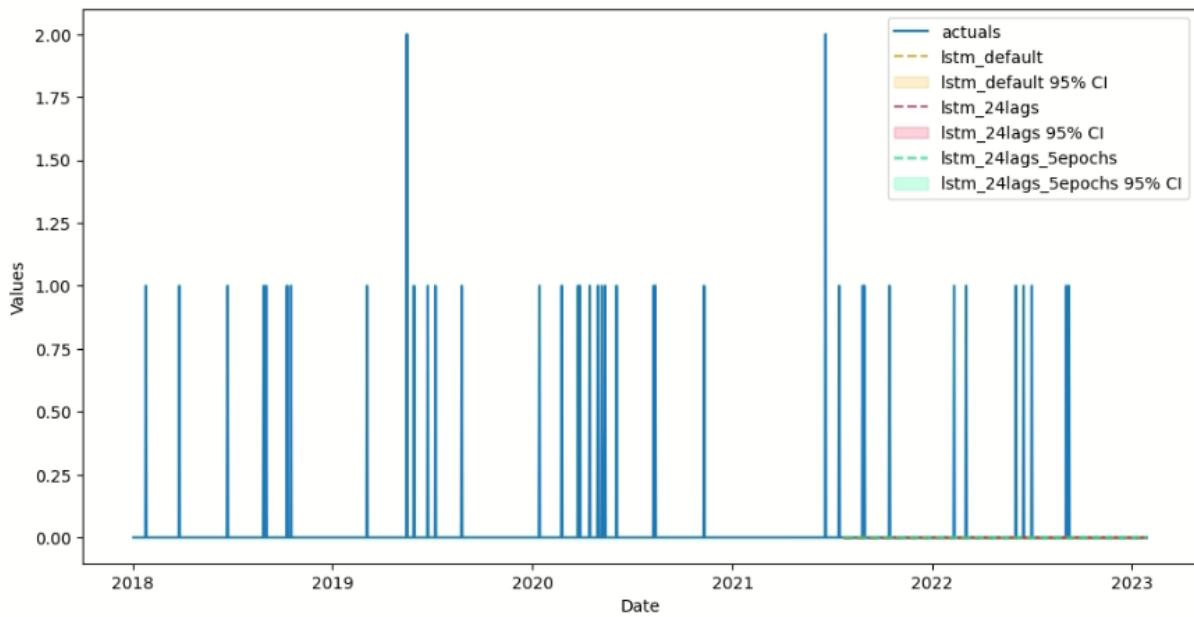


Como era de anticipar, este modelo inicial no proporcionó los resultados óptimos que se buscaban. No obstante, el hecho de que se hayan obtenido resultados de una manera tan directa constituye un excelente punto de partida. El desafío que presenta la refinación del modelo para que genere resultados más útiles, no debería ser particularmente abrumador.

Se iniciarán las modificaciones con ajustes relativamente simples, incrementando la cantidad de rezagos que el modelo considera para sus predicciones. El análisis PACF mostró una autocorrelación significativa de 24 meses, por lo que se considerará este factor en la estrategia de ajuste del modelo:

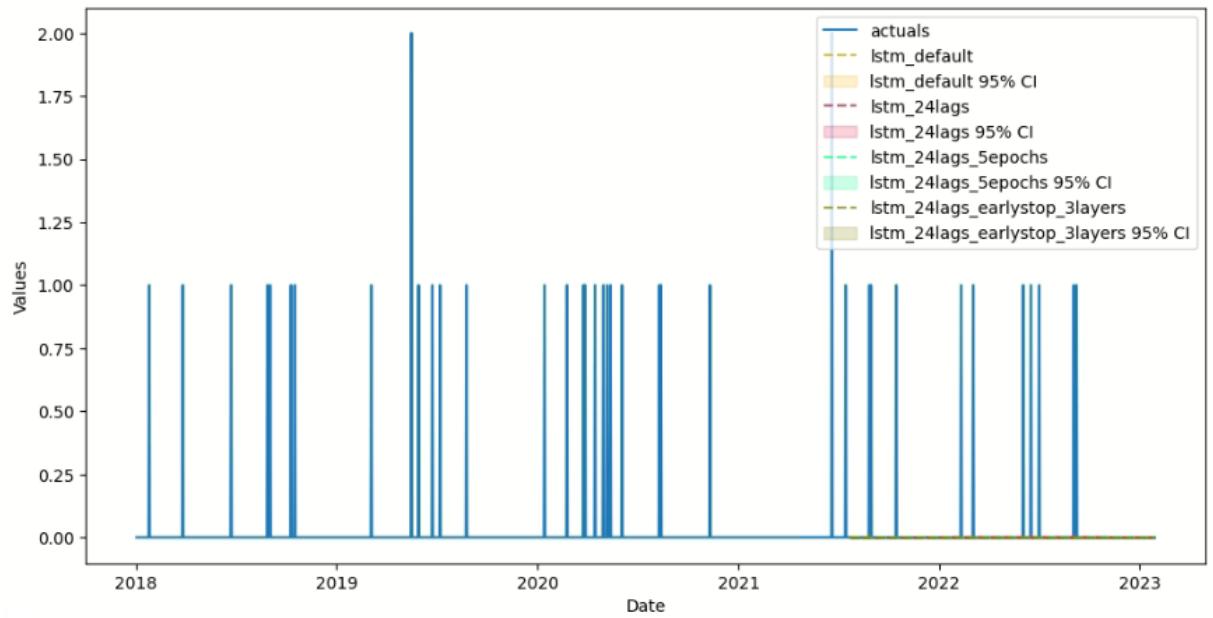


A partir de este punto, se observan algunos cambios que no son tan notables, sin embargo, aún se encuentra un camino considerable por recorrer hasta obtener un modelo completamente optimizado. Un siguiente paso razonable sería proporcionar un tiempo adicional para el entrenamiento. En este contexto, "tiempo adicional" se traduce en incrementar el número de épocas. Se explorará el efecto de aumentar a cinco el número de épocas. Adicionalmente, se implementará una validación del modelo en paralelo a su entrenamiento, especificando para ello un valor de validation split igual a 0.2 en la siguiente etapa:



Con cada iteración, el modelo avanza hacia resultados más resultados, pero todavía no está listo. En esta ocasión, algunos de los valores incluso caen dentro del intervalo de confianza del 95%. En el siguiente paso del análisis, se optará por incrementar la complejidad

de la red neuronal, aumentando el número de capas a tres y elevando el número de épocas a 25. No obstante, se mantendrá una monitorización constante del valor de la pérdida durante la validación. De esta forma, si no se observa mejora en este indicador tras cinco iteraciones, se le indicará al modelo que concluya el proceso de aprendizaje. Este mecanismo es comúnmente conocido como "early stopping" o detención temprana:



El progreso en el rendimiento del modelo sigue avanzando, aunque a un ritmo más gradual. Resulta comprensible si este proceso detallado de modelado puede parecer extenso. La prioridad recae en identificar un modelo que funcione eficientemente. Por lo tanto, se opta por avanzar directamente al mejor modelo que se logró identificar mediante este enfoque. A continuación, se presenta el código correspondiente:

```
# Define un callback para la detención temprana Mejor
early_stopping = EarlyStopping(monitor='val_loss', patience=10)

f.manual_forecast(
    call_me='lstm_best',
    lags=42,
    batch_size=32,
    epochs=1000,
    validation_split=.2,
    shuffle=True,
    activation='tanh',
    optimizer='Adam',
    learning_rate=0.005,
    lstm_layer_sizes=(72,) * 4,
    dropout=(0.2,) * 4, # Aumenta la tasa de dropout para promover la generalización
    callbacks=[early_stopping], # Utiliza la detención temprana para prevenir el sobreajuste
    plot_loss=True
)

f.plot_test_set(order_by='LevelTestSetMAPE', models='top_2', ci=True)
```

El código anterior representa la definición e implementación del modelo LSTM optimizado, que ha demostrado proporcionar resultados altamente satisfactorios. Este modelo incorpora varias mejoras y técnicas para promover la precisión y evitar el sobreajuste, lo que se refleja en la notable calidad de las gráficas de pérdida del modelo.

Primero, se establece un criterio para la detención temprana, que monitorea la 'val\_loss' y establece un límite de paciencia de 10 épocas sin mejoras para detener el entrenamiento. Este enfoque permite prevenir el sobreajuste y el gasto innecesario de recursos computacionales.

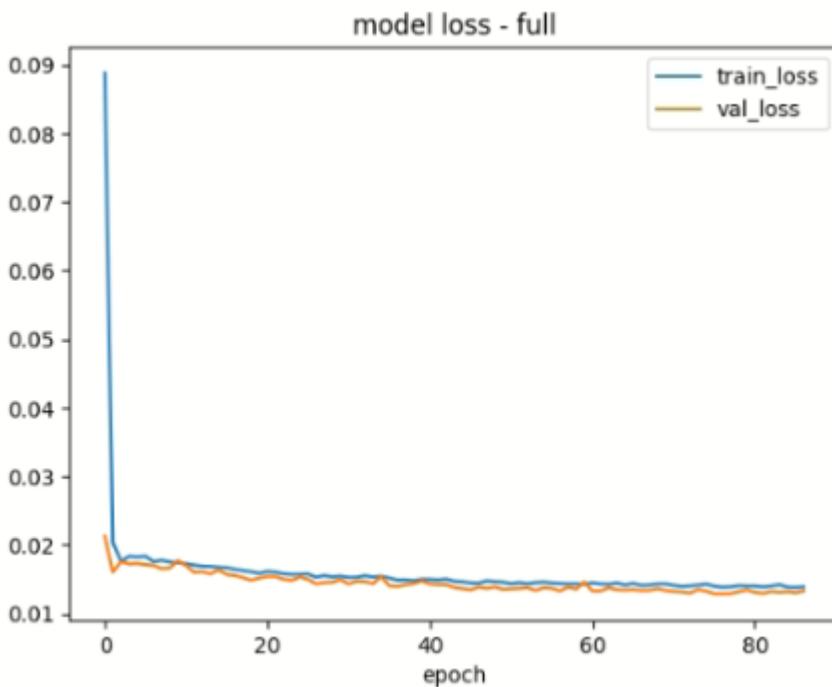
A continuación, se realiza el pronóstico manual utilizando el método 'manual\_forecast' del objeto 'Forecaster'. Este modelo toma en cuenta 42 rezagos, una mejora significativa en comparación con los modelos anteriores. El tamaño del lote se establece en 32, y el número de épocas se incrementa a 1000, proporcionando un margen amplio para que el modelo aprenda de los datos.

Se utiliza una división de validación del 20%, y los datos de entrenamiento se barajan antes de cada época para mejorar la generalización. La función de activación es 'tanh', y se utiliza el optimizador 'Adam' con una tasa de aprendizaje de 0.005.

El modelo LSTM está compuesto por cuatro capas, cada una con 72 unidades. Para ayudar a prevenir el sobreajuste, se incorpora un 'dropout' del 20% después de cada capa LSTM.

Por último, se incluye el callback de la detención temprana en el entrenamiento del modelo, y se grafica la pérdida del modelo durante el entrenamiento.

Los resultados muestran que el modelo ofrece un rendimiento de pérdida excepcional, con un perfil de pérdida suave y una tendencia descendente clara que indica que el modelo está aprendiendo efectivamente de los datos. Esto confirma que las mejoras y técnicas implementadas han llevado a una mejora sustancial en la calidad y rendimiento del modelo.



Ahora que finalmente se ha encontrado un modelo LSTM aceptable, es apropiado compararlo con un modelo sencillo, el más sencillo, la Regresión Lineal Múltiple (MLR). Como primer paso, se define un objeto "SeriesTransformer" que permitirá realizar transformaciones a los datos de la serie. En este caso, se aplica una diferenciación a la serie de tiempo para conseguir estabilidad, es decir, convertir los datos en una serie estacionaria. La diferenciación es un procedimiento común en el análisis de series temporales que ayuda a estabilizar la media de una serie eliminando los cambios en el nivel de una serie temporal, y por lo tanto, eliminando la tendencia y la estacionalidad.

El siguiente bloque de código añade diversos términos y regresores al modelo. Primero, se añaden 4 términos autorregresivos al modelo. Los términos autorregresivos permiten capturar las dependencias temporales en los datos. A continuación, se añaden regresores estacionales para el día, mes, trimestre y año. Estos regresores permiten capturar los patrones estacionales en los datos que se repiten en estos intervalos de tiempo. Finalmente, se añade un término de tendencia temporal para modelar cualquier tendencia subyacente en los datos.

```

transformer = SeriesTransformer(f)
f = transformer.DiffTransform()

f.add_ar_terms(4) # Añade términos autorregresivos para capturar dependencias diarias
f.add_seasonal_regressors('day', dummy=True) # Añade regresores estacionales diarios
f.add_seasonal_regressors('month','quarter', dummy=True) # Añade regresores estacionales mensuales
f.add_seasonal_regressors('year') # Añade regresores estacionales anuales
f.add_time_trend() # Añade un término de tendencia temporal
|

```

Después de configurar estos términos y regresores, se establece el estimador del modelo a 'mlr', lo que indica que se utilizará una regresión lineal múltiple para ajustar el modelo. Luego se realiza un pronóstico manual del modelo.

Finalmente, se revierte la transformación de diferenciación aplicada al inicio. Esto se hace para que los resultados del modelo puedan interpretarse en la escala original de los datos. En este paso, se excluyen todos los modelos LSTM previos de la reversión.

El último bloque de código genera un gráfico que muestra el desempeño de los dos modelos principales (el mejor modelo LSTM y el modelo MLR) en el conjunto de prueba. El desempeño se ordena por el error porcentual medio absoluto (MAPE) en el conjunto de pruebas.

```

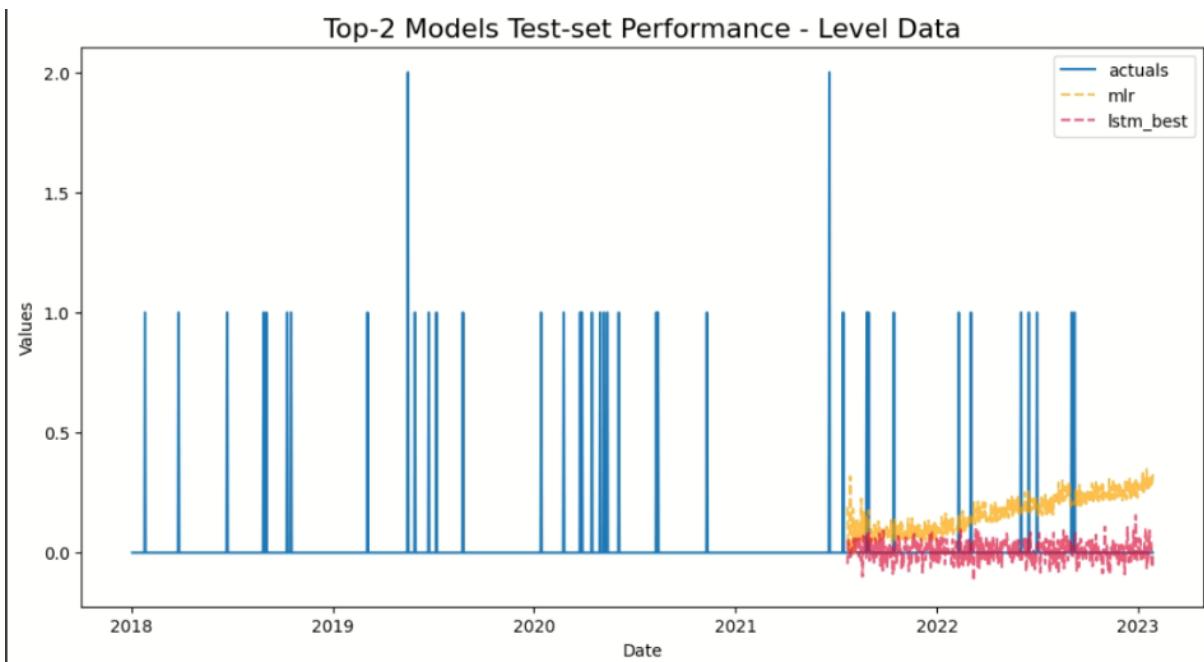
f.set_estimator('mlr')
f.manual_forecast(call_me='mlr')

f = transformer.DiffRevert(
    exclude_models = [m for m in f.history if m != 'mlr']
) # exclude all lstm models from the revert

f.plot_test_set(order_by='TestSetMAPE',models=['lstm_best','mlr'])
plt.title('Top-2 Models Test-set Performance - Level Data',size=16)
plt.show()

```

A pesar de los esfuerzos para mejorar y ajustar el modelo LSTM, los resultados obtenidos no se acercaron a los valores esperados. Por lo tanto, se tomó la decisión de no seguir utilizando el modelo LSTM para este análisis. Sin embargo, la exploración de modelos de series temporales no concluye aquí, y el siguiente paso será aplicar un enfoque multi parámetro, utilizando la regresión lineal múltiple, para la validación de los resultados.



### 5.2.7. Segunda fase de ajustes al diseño, implementación y validación al modelo planteado

Después de haber realizado enfoques con cGAN y LSTM para estudiar los feminicidios en Colombia, se decidió adoptar una nueva estrategia y cambiar los hiperparámetros para este tercer enfoque. Esta vez, el objetivo se centró en el análisis de los datos a nivel departamental, en lugar de nacional.

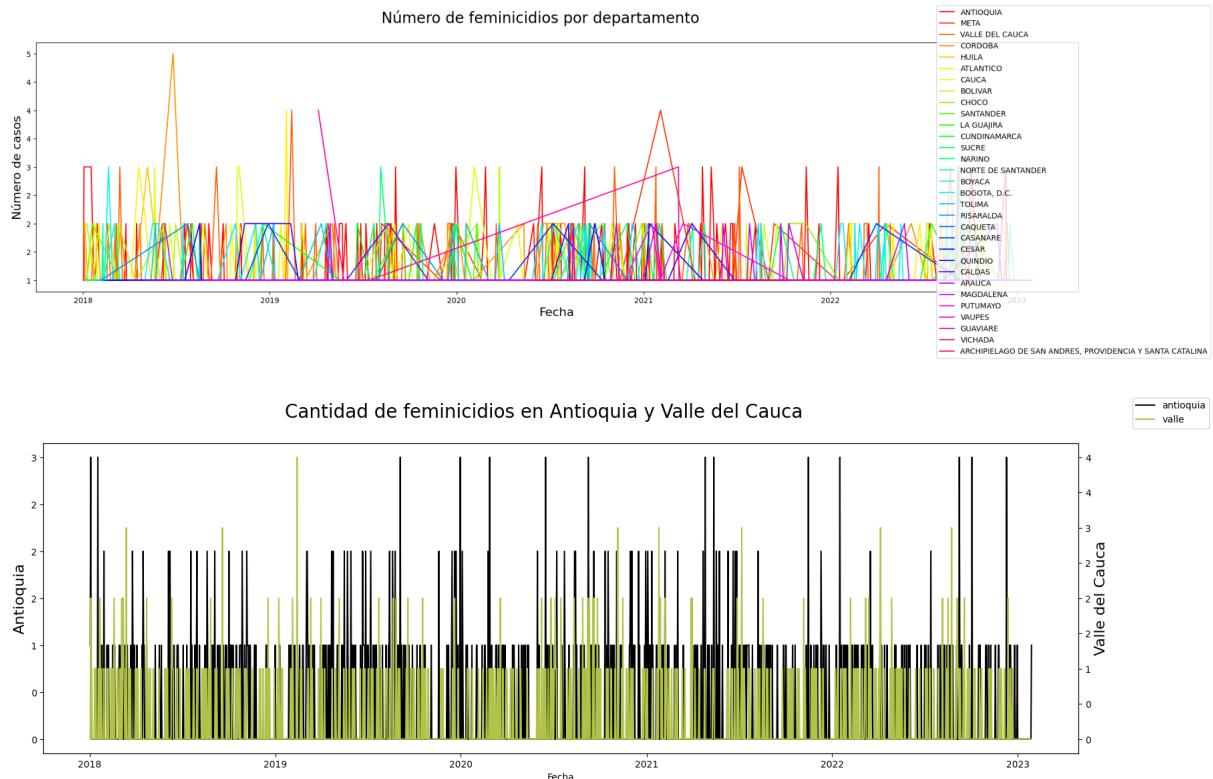
Se realizó una exploración inicial de los datos, donde se identificó que Antioquia y Valle del Cauca eran los dos departamentos con la mayor frecuencia de días con feminicidios registrados en el periodo de estudio, desde 2018 hasta 2023. Por lo tanto, se decidió centrar el análisis en estos dos departamentos.

Se organizó la información de tal manera que todas las fechas dentro del rango de estudio quedaran representadas, incluso aquellas sin casos de feminicidios, a las que se les asignó un valor de cero. De esta forma, se garantizó una continuidad en los datos.

Con los datos preparados, se optó por emplear una variedad de modelos de Machine Learning para generar pronósticos. Los modelos seleccionados incluyen Regresión Lineal Múltiple (mlr), ElasticNet, k-Nearest Neighbors (knn), Random Forest (rf), Gradient

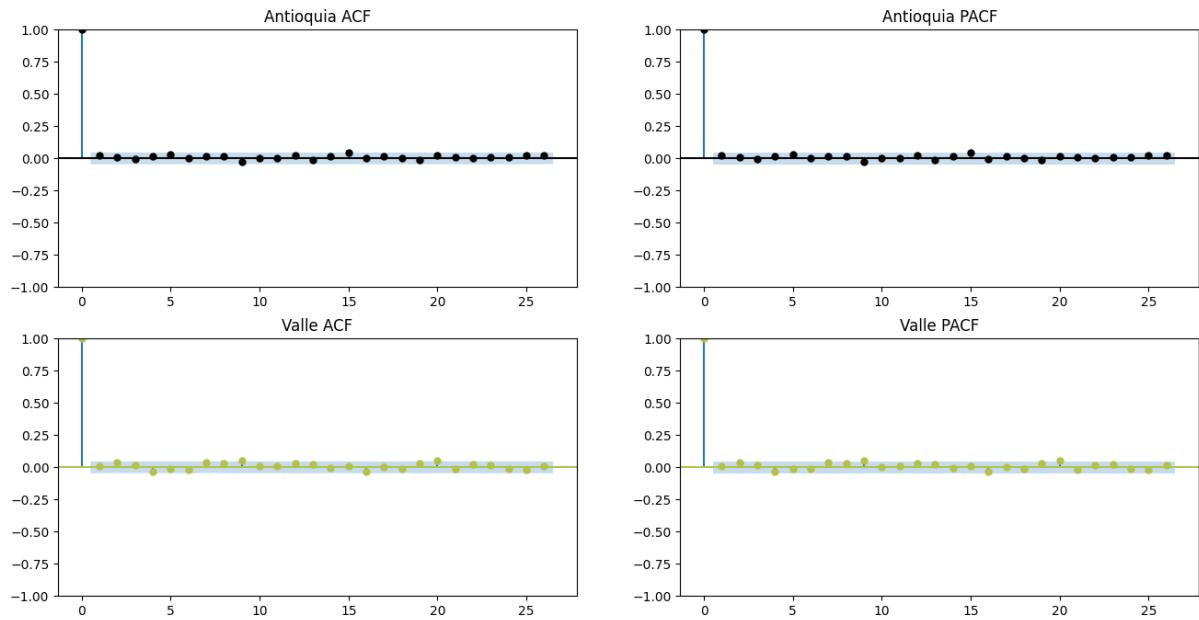
Boosting Trees (gbt), XGBoost y MultiLayer Perceptron (mlp). Para optimizar los hiperparámetros de cada modelo, se generaron cuadrículas utilizando el GridGenerator.

Se produjeron visualizaciones para comprender mejor las tendencias de los feminicidios en Antioquia y Valle del Cauca a lo largo del tiempo. Además, se calculó el coeficiente de correlación entre los feminicidios en ambos departamentos para determinar si existía alguna relación entre ellos.

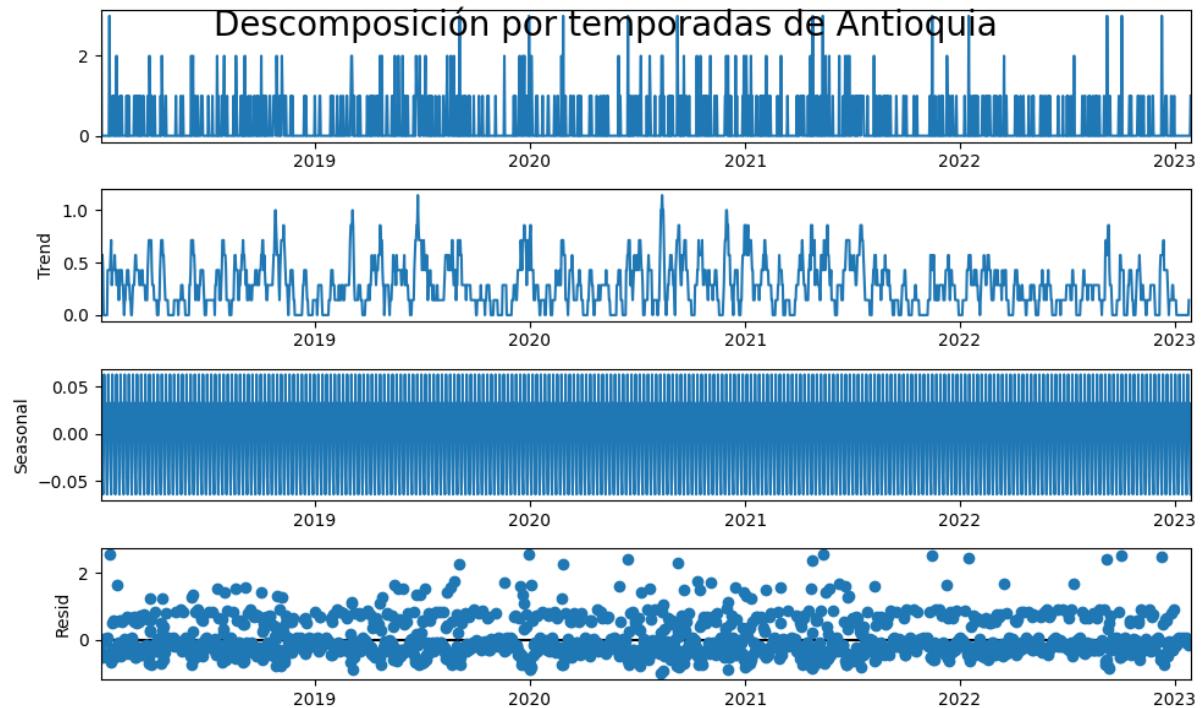


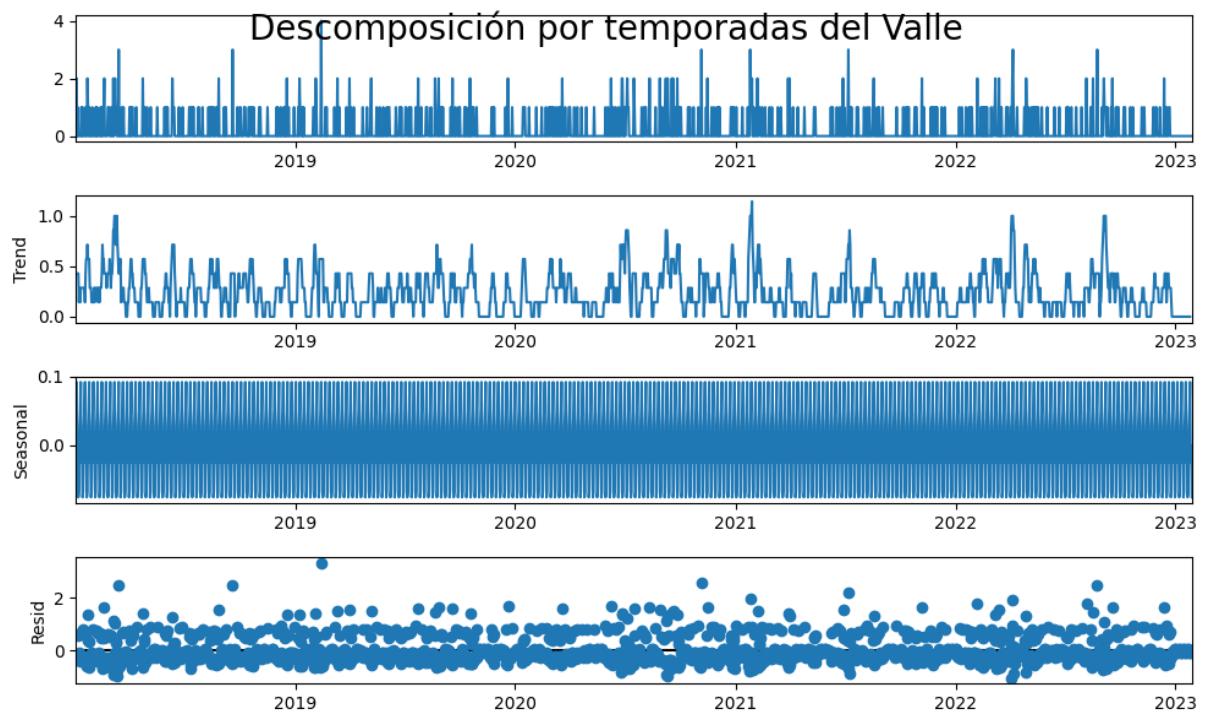
Finalmente, se crearon dos instancias de Forecaster para Antioquia y Valle del Cauca, definiendo la longitud de prueba, las fechas futuras, la longitud de la validación, las métricas y si se querían intervalos de confianza. Estos pronosticadores permitieron proporcionar predicciones útiles sobre los feminicidios futuros en cada departamento. Para lograrlo primero era necesario validar la estacionalidad de los datos. Para ello:

Primero, generé y visualicé las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) para ambos departamentos. Estos gráficos ayudan a entender la dependencia de una observación de las observaciones pasadas. En este caso, los lags que mostraron picos significativos en el ACF y PACF sugieren que estos pueden ser útiles para predecir el número de feminicidios.



Posteriormente, realicé una descomposición estacional de las series de tiempo de ambos departamentos. Esta técnica descompone una serie de tiempo en sus componentes de tendencia, estacionalidad y residuos. Las gráficas resultantes mostraron la tendencia subyacente de los feminicidios a lo largo del tiempo y la variabilidad estacional, si la hubiera.





Finalmente, apliqué el test de Augmented Dickey-Fuller (ADF) a las series de tiempo de Antioquia y Valle del Cauca. Esta prueba se utiliza para determinar si una serie de tiempo es estacionaria, es decir, si sus propiedades estadísticas, como la media y la varianza, son constantes a lo largo del tiempo.

```

Antioquia Augmented Dickey-Fuller results:
the test-stat value is: -42.16
the p-value is 0.0000
the series is stationary

-----
Valle del Cauca Augmented Dickey-Fuller results:
the test-stat value is: -29.25
the p-value is 0.0000
the series is stationary
-----
```

Los resultados del test de Augmented Dickey-Fuller para ambas series de tiempo mostraron un valor p de 0.0000, lo cual es menor que el valor crítico de 0.05. Esto indica que podemos rechazar la hipótesis nula de que las series de tiempo tienen una raíz unitaria y, por lo tanto, no son estacionarias.

Por lo tanto, concluimos que las series de tiempo de los feminicidios en Antioquia y Valle del Cauca son estacionarias. Esto es importante porque muchos modelos de series de tiempo asumen que los datos son estacionarios. Por lo tanto, el hecho de que nuestras series

de tiempo sean estacionarias es una buena señal para los siguientes pasos de nuestro análisis, que implicarán la aplicación de modelos de series de tiempo para pronosticar los futuros feminicidios en estos departamentos.

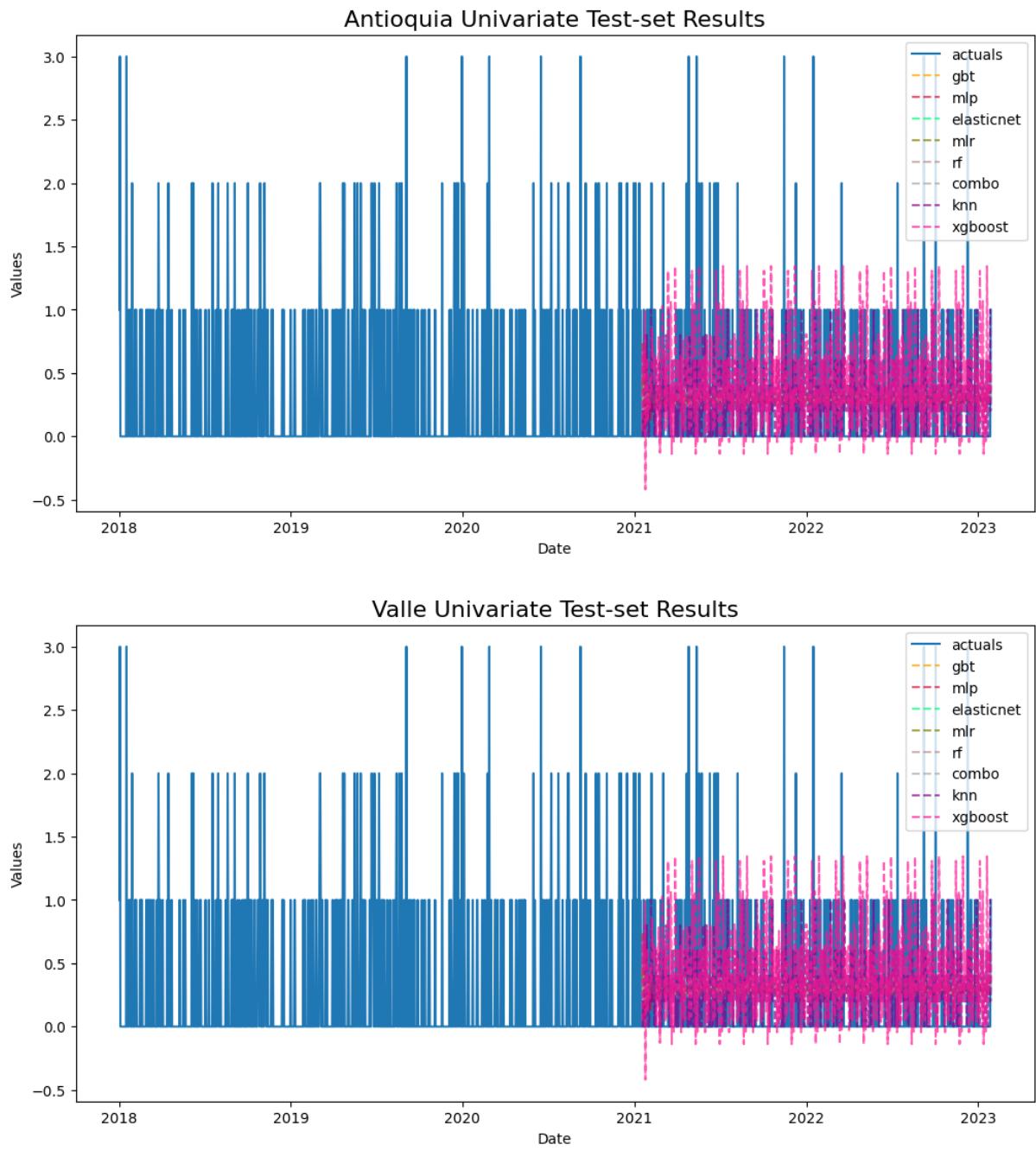
Una vez que se estableció la estacionariedad de las series, decidí implementar el enfoque univariado de Scalecast para pronosticar los feminicidios en Antioquia y Valle del Cauca.

Comencé seleccionando automáticamente las variables X para los modelos utilizando el estimador ElasticNet, probando hasta 50 valores de retardo y validando cruzadamente los resultados. Para Antioquia, el algoritmo seleccionó los 15 primeros valores de retardo (AR1 a AR15) como las variables más significativas. Para Valle del Cauca, el algoritmo seleccionó un mayor número de variables, hasta el valor de retardo 48 (AR1 a AR48).

Después de seleccionar las variables X, sintonicé y evalué varios modelos de aprendizaje automático, incluyendo Gradient Boosting Trees (gbt), Multilayer Perceptron (mlp), ElasticNet, Multiple Linear Regression (mlr), Random Forests (rf), k-nearest neighbors (knn), y XGBoost. También creé un modelo combinado (combo) que promedia las predicciones de los demás modelos.

Para cada modelo, calculé la importancia de las variables para entender cuáles eran las más importantes para predecir los feminicidios. También generé y visualicé las predicciones del conjunto de prueba para cada modelo, comparando su rendimiento con el de los valores reales.

Los resultados mostraron que, en general, los modelos tuvieron un rendimiento variable. Para Antioquia, el mejor modelo según el error cuadrático medio en el conjunto de prueba (TestSetRMSE) fue el gbt, mientras que para Valle del Cauca, el mejor modelo fue el combo.



Es interesante notar que a pesar de la variedad de técnicas de modelado utilizadas, no hubo un claro ganador entre los modelos. Esto sugiere que la tarea de predecir los feminicidios es compleja y que es probable que se necesiten enfoques más sofisticados o adicionales datos y variables para mejorar las predicciones.

Finalmente, resumí los resultados de los modelos en una tabla, mostrando las métricas clave para cada uno. Esta tabla proporciona una visión rápida y clara de la eficacia de cada modelo y ayudará a informar los próximos pasos de mi análisis.

ModelNickname	Series	TestSetRMSE	TestSetR2	InSampleRMSE	InSampleR2	best_model
0	gbt	Antioquia	0.5306	-0.0014	0.5305	0.0960
1	mlp	Antioquia	0.5323	-0.0079	0.5171	0.1410
2	elasticnet	Antioquia	0.5327	-0.0095	0.5579	0.0000
3	mlr	Antioquia	0.5333	-0.0116	0.5564	0.0056
4	rf	Antioquia	0.5372	-0.0265	0.5581	-0.0007
5	combo	Antioquia	0.5382	-0.0305	0.5189	0.1352
6	knn	Antioquia	0.5815	-0.2031	0.5119	0.1582
7	xgboost	Antioquia	0.6676	-0.5855	0.5256	0.1127
8	combo	Valle	0.5202	0.0048	0.4930	0.0825
9	mlp	Valle	0.5203	0.0046	0.5064	0.0317
10	mlr	Valle	0.5205	0.0037	0.5050	0.0373
11	gbt	Valle	0.5215	-0.0000	0.4895	0.0953
12	elasticnet	Valle	0.5216	-0.0003	0.5147	0.0000
13	rf	Valle	0.5220	-0.0021	0.5196	-0.0191
14	knn	Valle	0.5281	-0.0254	0.4968	0.0681
15	xgboost	Valle	0.5839	-0.2538	0.0967	0.9647

Para concluir la fase univariada de Scalecast-Forecast, calculé y mostré el error medio absoluto porcentual (MAPE) promedio de prueba y el coeficiente de determinación ( $R^2$ ) promedio de prueba para cada serie. Estas métricas nos proporcionan una medida resumida de la precisión y eficacia de los modelos que hemos construido.

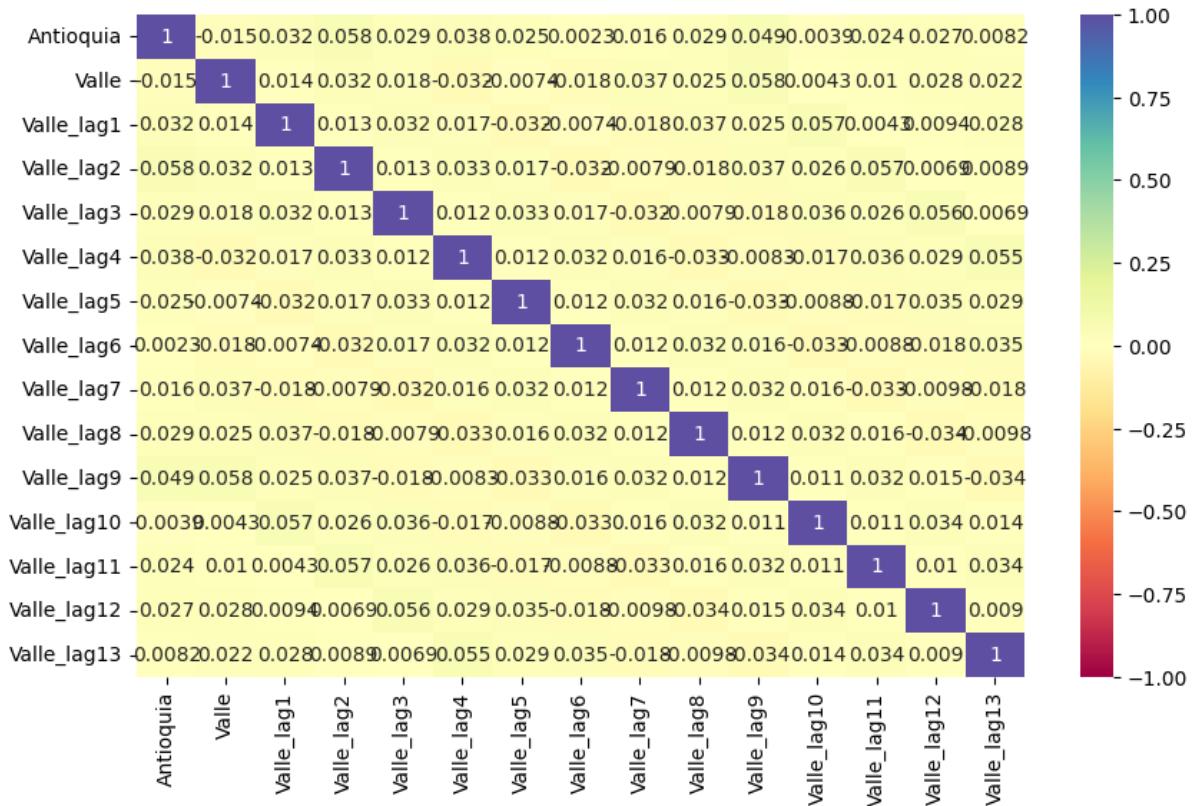
La MAPE promedio de prueba para Antioquia fue de 0.5567, mientras que para Valle del Cauca fue de 0.5298. Esto significa que, en promedio, los modelos se desviaron alrededor del 55.67% y el 52.98% de los valores reales para Antioquia y Valle del Cauca, respectivamente. Por otro lado, el  $R^2$  promedio de prueba fue de -0.11 para Antioquia y -0.03 para Valle del Cauca. Estos valores de  $R^2$  negativos sugieren que los modelos univariados no pudieron explicar eficientemente la variabilidad en los datos.

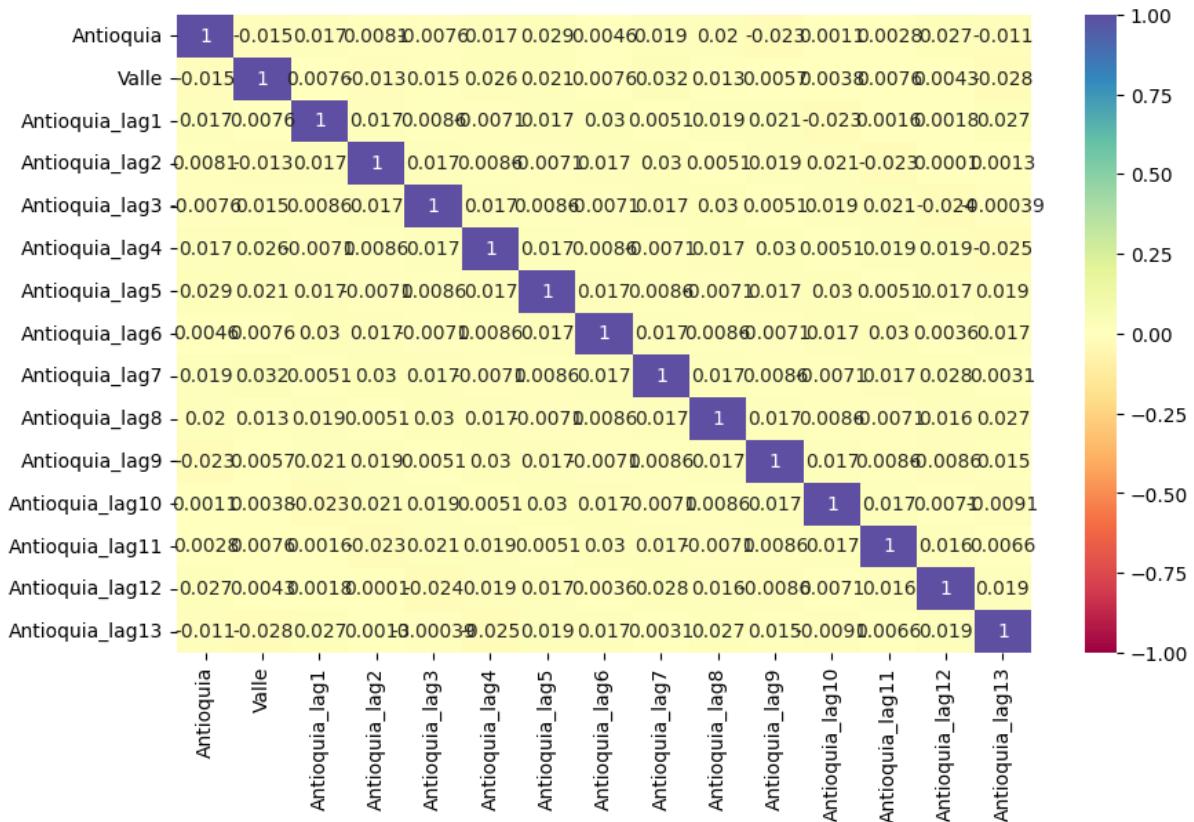
Con estos resultados en mente, decidí avanzar a la fase multivariada de Scalecast, donde combinaría y modelaría simultáneamente las series de tiempo de Antioquia y Valle del

Cauca. Inicialmente, creé un objeto MVForecaster para alojar las series y realizar los análisis necesarios.

Una parte crucial de la modelización multivariada es entender las relaciones entre las distintas series de tiempo. Por lo tanto, calculé y mostré las correlaciones entre las series y las correlaciones retardadas entre las series. Las correlaciones directas eran muy pequeñas, sugiriendo una relación débil entre las series. Sin embargo, las correlaciones retardadas pueden mostrar una imagen más detallada de cómo las series se influyen mutuamente con el tiempo.

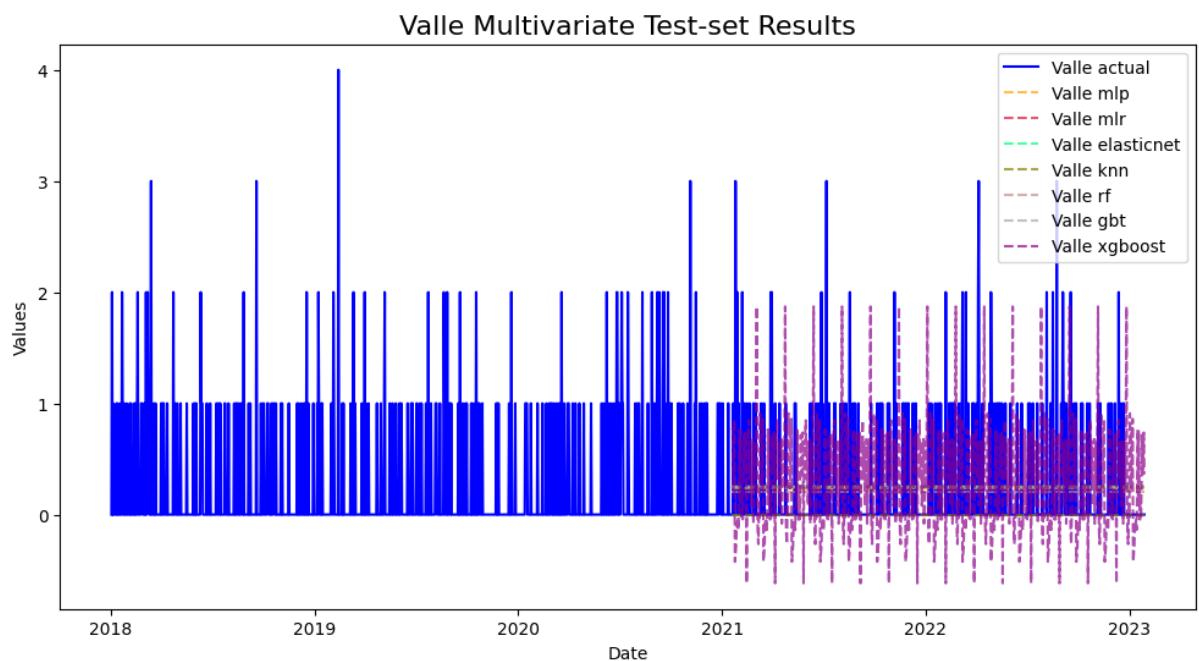
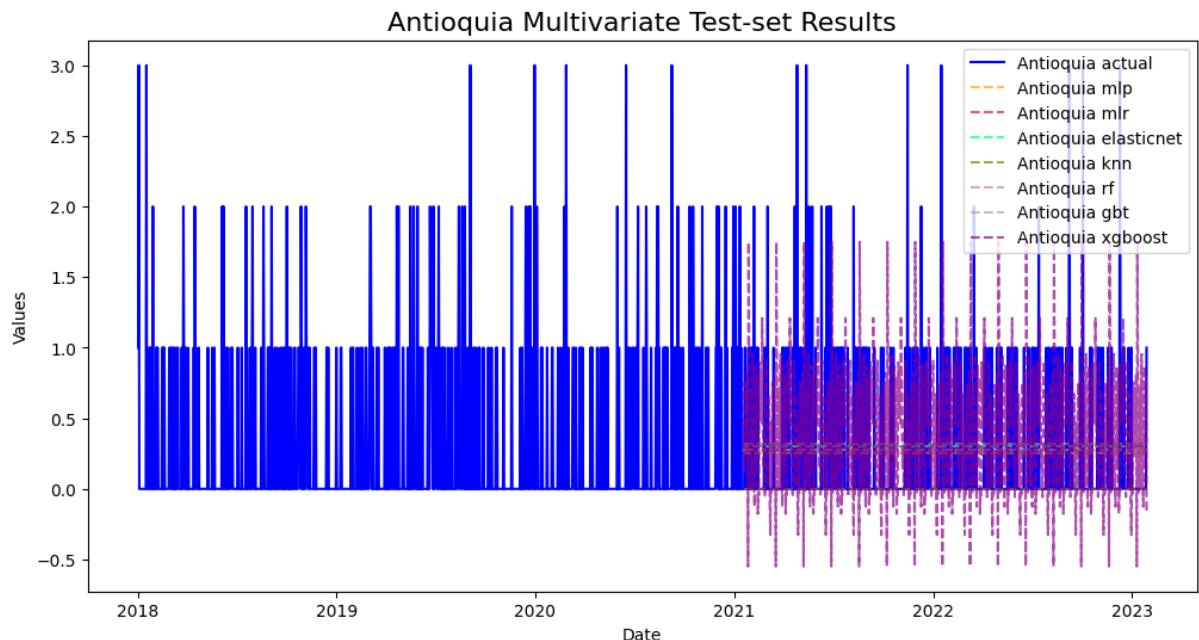
Generé y mostré mapas de calor para visualizar estas correlaciones retardadas, lo que proporciona una representación intuitiva de cómo las correlaciones cambian con diferentes retrasos. Estos gráficos y análisis son cruciales para entender las relaciones entre las series y guiarán la selección de variables para los modelos multivariados.





Después de configurar el optimizador para optimizar en la media, probé varios modelos en las series de Antioquia y Valle del Cauca y seleccioné el mejor modelo basándome en el RMSE del conjunto de prueba.

A continuación, se trazaron los resultados del conjunto de prueba multivariado para cada serie, lo que permitió visualizar el rendimiento de los diferentes modelos en los datos de prueba:



Finalmente, exporté un resumen de los modelos, que contiene detalles sobre el apodo del modelo, los parámetros del modelo, las métricas RMSE y R<sup>2</sup> para el conjunto de prueba y el conjunto de entrenamiento, y los retrasos utilizados

	ModelNickname	Series	HyperParams	TestSetRMSE	TestSetR2	InSampleRMSE	InSampleR2	Lags
0	mlp	Antioquia	{'activation': 'relu', 'hidden_layer_sizes': (25,), 'solver': 'adam', 'normalizer': 'minmax'}	0.5302	-0.0001	0.5557	0.0034	3
1	mlr	Antioquia	{'normalizer': None}	0.5323	-0.0081	0.5552	0.0052	3
2	elasticnet	Antioquia	{'alpha': 0.1, 'l1_ratio': 0, 'normalizer': 'scale'}	0.5324	-0.0081	0.5552	0.0051	3
3	knn	Antioquia	{'n_neighbors': 2}	0.5878	-0.2290	0.6131	-0.2135	3
4	rf	Antioquia	{'max_depth': 2, 'n_estimators': 100, 'max_features': 'sqrt', 'max_samples': 1}	0.5343	-0.0156	0.5678	-0.0282	1
5	gbt	Antioquia	{'max_depth': 2, 'max_features': 'sqrt'}	0.5309	-0.0025	0.5494	0.0257	3
6	xgboost	Antioquia	{'n_estimators': 150, 'scale_pos_weight': 10, 'learning_rate': 0.2, 'gamma': 0, 'subsample': 0.8}	0.7167	-0.8270	0.5195	0.1299	6
7	stacking	Antioquia	{'estimators': [('mlr', LinearRegression()), ('elasticnet', ElasticNet(alpha=0.1, l1_ratio=0)), ('mlp', MLPRegressor(hidden_layer_sizes=(25,))), 'final_estimator': KNeighborsRegressor(n_neighbors=2)}]	0.6766	-0.6286	0.6840	-0.5039	13
8	mlp	Valle	{'activation': 'relu', 'hidden_layer_sizes': (25,), 'solver': 'adam', 'normalizer': 'minmax'}	0.5037	-0.0005	0.5155	0.0027	3
9	mlr	Valle	{'normalizer': None}	0.5042	-0.0024	0.5157	0.0019	3
10	elasticnet	Valle	{'alpha': 0.1, 'l1_ratio': 0, 'normalizer': 'scale'}	0.5042	-0.0024	0.5157	0.0019	3
11	knn	Valle	{'n_neighbors': 2}	0.5523	-0.2027	0.5890	-0.3017	3
12	rf	Valle	{'max_depth': 2, 'n_estimators': 100, 'max_features': 'sqrt', 'max_samples': 1}	0.5043	-0.0028	0.5176	-0.0000	1
13	gbt	Valle	{'max_depth': 2, 'max_features': 'sqrt'}	0.5043	-0.0027	0.5099	0.0242	3
14	xgboost	Valle	{'n_estimators': 150, 'scale_pos_weight': 10, 'learning_rate': 0.2, 'gamma': 0, 'subsample': 0.8}	0.7148	-10.146	0.5016	0.0561	6
15	stacking	Valle	{'estimators': [('mlr', LinearRegression()), ('elasticnet', ElasticNet(alpha=0.1, l1_ratio=0)), ('mlp', MLPRegressor(hidden_layer_sizes=(25,))), 'final_estimator': KNeighborsRegressor(n_neighbors=2)}]	0.6223	-0.5273	0.6255	-0.4668	13

Para Antioquia, el modelo de perceptrón multicapa (MLP) con una función de activación ReLU, un tamaño de capa oculta de 25, y el algoritmo de optimización ADAM tuvo el menor RMSE en el conjunto de pruebas (0.5302), y un coeficiente de determinación  $R^2$  cercano a cero (-0.0001), lo que sugiere que el modelo no pudo explicar la variabilidad en los datos de Antioquia.

Para Valle del Cauca, el modelo MLP con las mismas configuraciones tuvo el menor RMSE en el conjunto de pruebas (0.5037), y un coeficiente de determinación  $R^2$  también cercano a cero (-0.0005), lo que sugiere que el modelo tampoco pudo explicar la variabilidad en los datos de Valle.

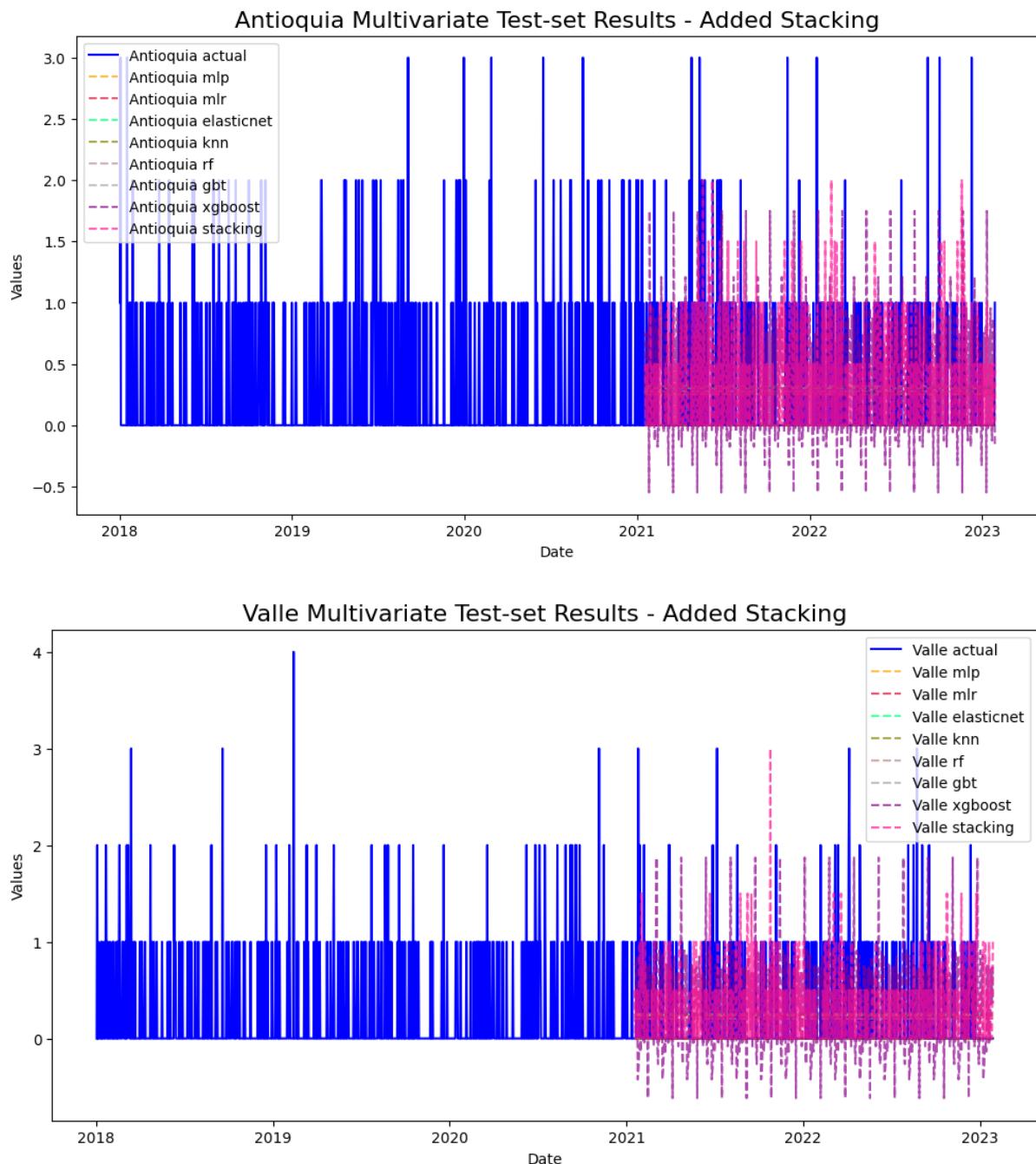
Estos resultados sugieren que los modelos multivariados no lograron desempeñarse mejor que los modelos univariados en esta tarea de pronóstico. Este hecho podría ser resultado de varias razones, como la falta de suficientes datos históricos, la falta de relaciones significativas entre las series de tiempo, o la necesidad de ajustar más finamente los hiperparámetros de los modelos.

Después de los resultados multivariados, se implementó un enfoque de apilamiento (o stacking). Esta es una técnica de aprendizaje supervisado en la que se combinan diferentes modelos de aprendizaje automático para mejorar el rendimiento predictivo. Los modelos que se utilizaron en el apilamiento incluyen la regresión lineal multivariable (mlr), la red elástica (elasticnet), el perceptrón multicapa (mlp), y el algoritmo de K-vecinos más cercanos (knn) como el estimador final.

Luego se realizó un pronóstico manual con 13 retrasos y se agregó el modelo de apilamiento al objeto del pronosticador multivariable.

Finalmente, se volvieron a trazar los resultados del conjunto de pruebas para las series de Antioquia y Valle, pero esta vez incluyendo los resultados del modelo de apilamiento.

Observando los resultados, se puede ver que el modelo MLP sigue siendo el que tiene el menor RMSE tanto para Antioquia como para Valle, siendo marcado como el "mejor modelo". El modelo de apilamiento no mejoró los resultados del conjunto de pruebas, obteniendo un RMSE de 0.6766 para Antioquia y 0.6223 para Valle, que son superiores a los RMSE de los otros modelos.



Estos resultados sugieren que, aunque el apilamiento puede ser útil en algunos casos, en esta instancia particular no logró mejorar las predicciones. Esto podría ser debido a una

variedad de factores, incluyendo la posibilidad de que los modelos individuales ya estuvieran dando buenos resultados, y la combinación de estos no aportó un beneficio adicional. También podría ser que los hiperparámetros del modelo de apilamiento necesiten un ajuste más detallado.

El backtesting es una técnica comúnmente utilizada en el análisis de series de tiempo para evaluar la precisión de un modelo predictivo. El proceso involucra el uso de datos históricos para generar predicciones y luego comparar estas predicciones con los resultados reales que ocurrieron durante ese período de tiempo.

En este caso, se usó backtesting para evaluar el desempeño de los modelos MLP (Perceptrón Multicapa) en las series de tiempo de Antioquia y Valle. El backtesting se llevó a cabo en tres iteraciones, con una longitud de pronóstico de 52 y saltando atrás 4 pasos de tiempo después de cada iteración.

Los resultados muestran el RMSE (Root Mean Squared Error) y R<sup>2</sup> (coeficiente de determinación) para cada iteración y las medias de estos valores.

			Iter0	Iter1	Iter2	Average
Series	Model	Metric				
Antioquia	mlp	rmse	0.5249	0.5672	0.5505	0.5475
		r2	-0.0177	-0.2443	-0.1724	-0.1448
Valle	mlp	rmse	0.3635	0.4338	0.4004	0.3992
		r2	-0.0542	-0.3390	-0.0345	-0.1426

Para Antioquia, el RMSE promedio es de 0.5475 y el R<sup>2</sup> promedio es de -0.1448. Los valores de RMSE más bajos son mejores, ya que indican un menor error en las predicciones, y los valores de R<sup>2</sup> más altos son mejores, ya que indican una mayor proporción de la variación total de los datos explicada por el modelo. El R<sup>2</sup> negativo sugiere que el modelo no explica la variabilidad de la respuesta en torno a su media.

Para Valle, el RMSE promedio es de 0.3992 y el R<sup>2</sup> promedio es de -0.1426. En este caso, también se ve que el modelo no tiene una buena precisión de predicción según los indicadores de R<sup>2</sup>.

Estos resultados sugieren que, aunque el modelo MLP fue seleccionado como el "mejor modelo" en función del RMSE en el conjunto de pruebas, su rendimiento en el backtesting no es óptimo. Esto puede deberse a una variedad de factores, incluyendo cambios en las series de tiempo que no se reflejan en los datos de entrenamiento, o simplemente que el modelo no es adecuado para estas series de tiempo. También podría ser necesario un ajuste más detallado de los hiper parámetros del modelo.

### 5.3. Aportes:

El presente proyecto tiene como objetivo principal explorar la viabilidad y eficacia de las Generative Adversarial Networks condicionales (cGAN) en la tarea de predicción de feminicidios. A pesar de las limitaciones encontradas en términos de datos y recursos computacionales, se espera que el estudio pueda aportar valiosos conocimientos en las siguientes áreas:

- **Mejora de la comprensión sobre la aplicación de cGAN en la predicción de eventos sociales:** El estudio explorará el uso de cGAN, una técnica de machine learning avanzada, en un contexto social y criminológico. Esto puede aportar conocimientos valiosos sobre cómo las técnicas de inteligencia artificial pueden utilizarse en la predicción y, en última instancia, en la prevención de crímenes graves como los feminicidios.
- **Desarrollo de metodologías para superar las limitaciones de los recursos computacionales y de datos:** Debido a las restricciones de recursos y datos, el proyecto también buscará desarrollar estrategias y enfoques efectivos para superar estos desafíos. Esto puede ser útil para otros investigadores y profesionales que también estén trabajando en entornos de recursos limitados.
- **Contribución al campo de la criminología y la prevención del crimen:** Al proporcionar un modelo predictivo de feminicidios, se espera que este proyecto pueda contribuir de manera significativa al campo de la criminología y la prevención del crimen. El modelo podría ser potencialmente útil para los encargados de formular políticas y los profesionales en la planificación e implementación de estrategias de prevención del feminicidio.

El éxito de este proyecto depende de la colaboración y contribución de todos los participantes. Cada miembro del equipo, Juan Jose Restrepo Bonilla, Juan Camilo Tobar, Christian Camilo Urcuqui, Sergio Alejandro Cabrera y Jhon Jairo Restrepo, aportará su experiencia y habilidades únicas para superar los desafíos y lograr los objetivos del proyecto.

#### 5.4. Entregables:

- **Modelo cGAN No Entrenado y Justificación:** Debido a las limitaciones en los recursos de cómputo y la naturaleza intensiva del modelo cGAN, no fue posible realizar su entrenamiento. Sin embargo, se proporcionará la implementación inicial del modelo cGAN, así como una discusión detallada sobre los desafíos y las limitaciones que impidieron su entrenamiento. Este entregable servirá como base para futuras investigaciones, una vez que se disponga de más recursos de computación, y ayudará a otros a comprender los desafíos inherentes al entrenamiento de modelos cGAN.
- **Código del Proyecto:** Este entregable incluirá todos los scripts y archivos de código fuente utilizados en el proyecto. Esto abarca el código de preprocesamiento de datos, los scripts de visualización de datos y cualquier otro código relevante.
- **Documentación del Proyecto:** Este entregable proporcionará una descripción detallada del proyecto, los métodos utilizados, las dificultades encontradas y las decisiones tomadas. La documentación incluirá tanto aspectos técnicos (por ejemplo, explicaciones de cómo se utilizó el código, la elección de parámetros del modelo, etc.) como la guía del usuario final (por ejemplo, cómo utilizar el dashboard de BI, cómo interpretar los resultados, etc.)
- **Datos:** Este entregable consistirá en el conjunto de datos utilizados en el proyecto. Dado que los datos incluyen información sensible, como las fechas de los crímenes de feminicidio, se manejarán con el debido cuidado y consideración. Dependiendo de las normativas, se puede proporcionar acceso a los datos procesados que se utilizaron en el análisis.
- **Dashboard de BI:** Este entregable es un producto de un concurso de Datajam y proporcionará una visualización interactiva de los datos. El dashboard incluirá gráficos de los datos de entrada y cualquier otro análisis descriptivo que se haya llevado a cabo.

- **Informe Final:** Este entregable será un informe completo que detallará todos los aspectos del proyecto, incluyendo la motivación del proyecto, la descripción detallada de los métodos y técnicas utilizadas, los desafíos encontrados y cómo se superaron, y las recomendaciones para futuros trabajos.
- **Presentación del Proyecto:** Finalmente, se proporcionará una presentación que resuma el proyecto y sus resultados de manera concisa y visualmente atractiva, apta para presentar a una audiencia general o a posibles interesados.

### 5.5. Conclusiones:

Con base en las actividades realizadas y los resultados obtenidos durante el desarrollo del proyecto, a continuación se presentan las conclusiones correspondientes:

- **Limitaciones del Modelo cGAN:** El proyecto evidenció que, a pesar del potencial de los modelos cGAN para la generación de datos realistas y la predicción de eventos, su entrenamiento requiere una cantidad significativa de recursos computacionales. En este caso, la falta de disponibilidad de tales recursos impidió que se pudiera entrenar el modelo cGAN.**Exploración de Alternativas de Predicción:** Dada la imposibilidad de entrenar el modelo cGAN, se exploraron alternativas viables para la predicción de feminicidios, específicamente métodos de Forecasting Multivariate y Forecasting Univariate. Estas técnicas, específicamente el método de Forecasting Multivariate, resultaron ser más factibles dada la limitación de recursos y demostraron ser útiles para la tarea. El método de Forecasting Multivariate fue particularmente efectivo, logrando una mayor proximidad a las predicciones reales en comparación con el método de Forecasting Univariate.
- **Importancia de la Colaboración para la Obtención de Datos:** El éxito del proyecto depende en gran medida de la colaboración con el Observatorio de Feminicidios, entidad que facilitó un conjunto crucial de datos de casos de feminicidios. Este conjunto de datos permitió la formación y validación de los modelos predictivos, siendo la piedra angular de todo el proceso de análisis. Esta colaboración subrayó la importancia de establecer fuertes relaciones con entidades y organizaciones que pueden proporcionar acceso a datos vitales y relevantes.

Sin embargo, trabajar con estos datos también presentó desafíos únicos. Debido a la naturaleza sensible de la información, fue necesario manejar los datos con sumo cuidado, garantizando el respeto por la privacidad y la confidencialidad. A través de este proceso, se reforzó la comprensión del equipo sobre la importancia de la ética y la privacidad en la ciencia de datos, especialmente al tratar con temas tan delicados y personales.

En general, la experiencia de colaboración con el Observatorio de Feminicidios demostró ser invaluable, no sólo proporcionando los datos necesarios para el proyecto, sino también ofreciendo una visión única y valiosa sobre la importancia y el impacto del trabajo que se estaba realizando

En conclusión, a pesar de los desafíos, el proyecto logró resultados importantes y proporcionó valiosas lecciones y experiencias. Estos hallazgos pueden ser de gran ayuda para otros que busquen aplicar técnicas de machine learning a problemas similares, y contribuyen significativamente a la base de conocimientos existente sobre la predicción de feminicidios y la aplicación del machine learning a problemas del mundo real.

## 5.6. Trabajo Futuro

El trabajo futuro para este proyecto se centra principalmente en superar las limitaciones encontradas durante su desarrollo y en aprovechar los resultados obtenidos hasta el momento. A continuación, se detallan algunas de las principales recomendaciones para futuros trabajos en esta línea de investigación:

- **Entrenamiento del modelo cGAN:** Con la disponibilidad de recursos de cómputo más potentes, será interesante intentar de nuevo el entrenamiento del modelo cGAN, utilizando los datos de feminicidios. El potencial de las cGAN para generar datos realistas y precisos podría proporcionar resultados importantes para la predicción y el análisis de feminicidios.
- **Ampliación y mejoramiento de los datos:** La calidad y cantidad de datos disponibles influyen significativamente en los resultados de cualquier modelo de machine learning. Para futuros trabajos, sería útil buscar formas de ampliar y mejorar los datos utilizados en este proyecto. Esto podría implicar la inclusión de más datos históricos,

la recopilación de datos más granulares o la inclusión de variables adicionales que podrían ser relevantes para la predicción de feminicidios.

- **Exploración de otros modelos y técnicas:** Aunque el forecasting multivariate y univariate mostraron ser técnicas útiles en este caso, existen muchos otros modelos y técnicas que podrían ser aplicables para la predicción de feminicidios. Futuras investigaciones podrían explorar el uso de otros enfoques, como el machine learning basado en secuencias (como las redes neuronales recurrentes), los modelos de regresión, entre otros.
- **Implementación de un sistema de predicción en tiempo real:** Con base en los modelos desarrollados y los resultados obtenidos, se podría trabajar en el diseño e implementación de un sistema de predicción en tiempo real de feminicidios. Este sistema podría ser de gran utilidad para las autoridades y las organizaciones que trabajan en la prevención de este tipo de delitos.
- **Evaluación del impacto de las predicciones:** Una vez implementado un sistema de predicción, será crucial evaluar su impacto y efectividad. Esto podría incluir el análisis de cómo las predicciones se traducen en acciones preventivas y de qué manera contribuyen a la disminución de los feminicidios.
- Estas recomendaciones ofrecen una ruta prometedora para la continua investigación y desarrollo en el campo de la predicción de feminicidios, y tienen el potencial de contribuir significativamente a la prevención de estos crímenes en el futuro.

## 6. Referencias bibliográficas

1. Amarante, V. (2015). La medición de la desigualdad: otros indicadores. *Naciones Unidas*, 12.
2. Diaz, A. (2022, 9 de julio). *Medición, detección y reducción de algoritmos de IA*. damaspain. <https://www.damaspain.org/medicion-sesgos-algoritmos/>
3. Silberg, J. y Manyika, J. (2019, 6 de junio). *Tackling bias in artificial*

*intelligence (and in humans).*  
McKinsey&Company.<https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans>.

4. Adeyanju, O. (2020, 19 de junio). *How Decentralization Could Alleviate Data Biases In Artificial Intelligence.* Forbes.<https://www.forbes.com/sites/oluwaseunadeyanju/2020/06/19/how-decentralization-could-alleviate-data-biases-in-artificial-intelligence/?sh=22cc69005248>.
5. *Colombia tiene 1.400 obras inconclusas, por un total de \$25 billones de pesos.* (2020, 10 de septiembre). ElQuindiano.<https://www.elquindiano.com/noticia/21226/colombia-tiene-1400-obras-inconclusas-por-un-total-de-25-billones-de-pesos>.
6. BOHÓRQUEZ GUEVARA, K. S. (2020, 3 de abril). *Hay 1.901 proyectos de vivienda paralizados en el país por la emergencia del Covid-19.* La Republica.  
<https://www.larepublica.co/economia/hay-1-901-proyectos-de-vivienda-paralizados-en-el-pais-por-la-emergencia-del-covid-19-2987326>.
7. Valderrama Rojas, C. L., Durán Pabón, I. M., Arboleda Urrego, O. J., Montealegre Arboleda, I. y Espinosa, S. (2022). Guía de Referencia para la adopción e implementación de proyectos con tecnología blockchain para el Estado Colombiano. *Ministerio de las TIC*, (2), 82.
8. Salgado, R. (2022, 20 mayo). *Principales amenazas que pueden dinamitar un proyecto de IA.* Think Big.  
<https://empresas.blogthinkbig.com/amenazas-para-proyectos-de-inteligencia-artificial/>

9. Ramírez Arce, J. L. (2018). El Feminicidio en Colombia. (Trabajo de Especialización). Universidad La Gran Colombia, Facultad de Postgrados, Especialización en Derecho Penal y Criminología. Dirigido por A. Díaz Pedrozo.
10. Bernal Aldana, A. N. (2020). El feminicidio: una revisión histórica del concepto. (Trabajo de Especialización). Universidad Santo Tomás, Tunja.
11. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems.
12. Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition.
13. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A. P., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition.
14. Iizuka, S., Simo-Serra, E., & Ishikawa, H. (2017). Globally and locally consistent image completion. ACM Transactions on Graphics (TOG).
15. Na8. (2019). Pronóstico de Series Temporales con Redes Neuronales en Python. Aprende Machine Learning. Recuperado de <https://www.aprendemachinelearning.com/pronostico-de-series-temporales-con-redes-neuronales-en-python/>

16. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., ... & Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Disponible en <https://www.tensorflow.org/about/bib>
  
  
  
17. Fundación Feminicidios Colombia. (2020). No somos un día: Informe Anual 2020. <https://www.feminicidioscolombia.org>
  
  
  
18. Hernández Severiche, M. J., & Díaz-Rincón, S. V. (2021). El feminicidio dentro del contexto jurídico-penal colombiano y su incidencia en la sociedad. Facultad de Ciencias Jurídicas y Sociales, Universidad Simón Bolívar.
19. Olah, C. (2015). Understanding LSTM Networks. Recuperado de <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
20. Pokharna, H. (2016). The best explanation of Convolutional Neural Networks on the internet. Recuperado de <https://medium.com/technologymadeeasy/the-best-explanation-of-convolutional-neural-networks-on-the-internet-fbb8b1ad5df8>
21. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).
22. Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.
23. Oli, N. (2018, April 2). An intuitive introduction to Generative Adversarial Networks (GANs). Medium. <https://medium.com/free-code-camp/an-intuitive-introduction-to-generative-adversarial-networks-gans-7a2264a81394>
24. Mirza, M., & Osindero, S. (2014). Conditional Generative Adversarial Nets. ArXiv Preprint. <https://arxiv.org/abs/1411.1784>.
25. OECD. (2021). OECD.AI Policy Observatory: Colombia. [En línea]. Disponible en:

<https://oecd.ai/dashboards/policy-initiatives/740/marco-etico-para-el-desarrollo-y-la-implementacion-de-la-inteligencia-artificial-en-colombia>

26. Guío Español, A., Tamayo Uribe, E., Gómez Ayerbe, P., & Mujica, M. P. (2021). Marco Ético para la Inteligencia Artificial en Colombia. Departamento Administrativo de la Presidencia de la República. [https://inteligenciaartificial.gov.co/static/img/MARCO\\_ETICO.pdf](https://inteligenciaartificial.gov.co/static/img/MARCO_ETICO.pdf)
27. ACNUDH (2021). Los derechos humanos y las nuevas tecnologías: la inteligencia artificial. Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos.
28. Stats New Zealand (2018). Algorithm Assessment Report. Gobierno de Nueva Zelanda.
29. Gimenez, G. (2018). Data Cleansing: 6 Steps to Clean Data. Datafloq.
30. Doshi-Velez, F., & Kortz, M. (2017). Accountability of AI Under the Law: The Role of Explanation. arXiv preprint arXiv:1711.01134.
31. World Wide Web Foundation. (2018). Artificial Intelligence - The Road Ahead in Low and Middle-Income Countries. World Wide Web Foundation.
32. IMDA & PDPC. (2020). Model Artificial Intelligence Governance Framework Second Edition. <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf>
33. IBM. (2021.). Inteligencia artificial (IA). <https://www.ibm.com/mx-es/topics/artificial-intelligence>
34. Amazon. (2022). ¿Qué es el aprendizaje profundo? <https://aws.amazon.com/es/what-is/deep-learning/>
35. Andrej Karpathy, & Fei-Fei Li. (n.d.). Convolutional Neural Networks (CNNs / ConvNets). CS231n Convolutional Neural Networks for Visual Recognition. Retrieved May 24, 2023, from <https://cs231n.github.io/convolutional-networks/#overview>

## 7. Anexos

### 7.1. Forecasting Multivariate:

<https://colab.research.google.com/drive/1SlyUY38yEaQp-1d3BbuqjD42fpnLTkw5#scrollTo=ZG88s8xirkeF>

- 7.2. Reporte:
- 7.3.  Reporte de Cumplimiento y Uso de las Herramientas de Principios Éticos
- 7.4. cGan model:  
<https://colab.research.google.com/drive/1XNVX-LnL7Exx6Z7JTNApbsHGNuROdUo?usp=sharing>
- 7.5. Dashboard de bi:  
<https://lookerstudio.google.com/reporting/5733ab5d-6f91-4f94-af9a-08bd39c44be6>
- 7.6. Forecasting Univariate:  
<https://drive.google.com/file/d/1wXYSgWuYtMSVXcHgWYg9K54lvu7dWqZo/view?usp=sharing>