IUS ET SCIENTIA
2024 • Vol. 10 • Nº 1 • ISSN 2444-8478
https://editorial.us.es/es/revistas/ius-et-scientia
https://dx.doi.org/10.12795/IETSCIENTIA • © Editorial Universidad de Sevilla 2024

Revelando lo no reportado: extracción de eventos basada en IA para analizar la representación estadounidense de los delitos de odio

UNVEILING THE UNREPORTED: AI-BASED EVENT EXTRACTION FOR ANALYZING THE AMERICAN REPRESENTATION OF HATE CRIMES

Daniel Suárez Alonso
Universidad Europea Miguel Cervantes
dsuarez@uemc.es 0000-0002-4505-2942

Recibido: 16 de abril de 2024 | Aceptado: 30 de mayo de 2024

RESUMEN

Los informes oficiales de delitos de odio en los Estados Unidos están subestimados en comparación con la cantidad real de incidentes de este tipo. Además, a pesar de las aproximaciones estadísticas, no hay informes oficiales de muchas ciudades estadounidenses sobre incidentes de odio. Aquí, mostramos inicialmente que la extracción de eventos y el aprendizaje multi-instancia, basados en inteligencia artificial (IA), aplicados a un conjunto de artículos de noticias locales, pueden predecir casos de delitos de odio. Luego utilizamos el modelo entrenado de IA para detectar incidentes de odio en ciudades para las cuales el FBI carece de estadísticas. Finalmente, entrenamos modelos de IA para predecir homicidios y secuestros, comparamos las predicciones con los informes del FBI y establecemos que, de hecho, los incidentes de odio están subestimados en comparación con otros tipos de delitos en la prensa local. Es importante destacar que esta información no ha sido extraída de este lugar.

PALABRAS CLAVE

Delitos de odio
Informes oficiales
Aproximaciones estadísticas
Extracción de eventos
Aprendizaje multi-instancia
Inteligencia artificial (IA)

ABSTRACT

Official reports of hate crimes in the United States are underestimated compared to the actual number of such incidents. Additionally, despite statistical approximations, many American cities lack official reports on hate incidents. Here, we initially demonstrate that event extraction and multi-instance learning, based on artificial intelligence (AI), applied to a set of local news articles, can predict hate crime cases. We then use the AI-trained model to detect hate incidents in cities for which the FBI lacks Official reports of hate crimes in the United States are underestimated compared to the actual number of such incidents. Additionally, despite

KEYWORDS

Hate crimes
Official reports
Statistical approximations
Event extraction
Multi-instance learning
Artificial intelligence (AI)

statistical approximations, many American cities lack official reports on hate incidents. Here, we initially demonstrate that event extraction and multi-instance learning, based on artificial intelligence (AI), applied to a set of local news articles, can predict hate crime cases. We then use the AI-trained model to detect hate incidents in cities for which the FBI lacks

I. Introducción

Los delitos de odio han adquirido un gran protagonismo en estos últimos años, ocupando un puesto destacado en las agendas de los Estados para su estudio y prevención (Aba Catoira, 2015, 42). Esto puede deberse a varios factores como por ejemplo a un aumento en los últimos años (Quesada Alcalá, 2015, 16), por una mejora en el registro de datos estadísticos por parte de las autoridades competentes (González Gaya, Domingo Navas, & Sebastián Perez, 2013, 114) o por una mayor confianza en interponer denuncia ante las Fuerzas y Cuerpos de Seguridad (Mercader, 2018, 21). También podría deberse a grandes cambios que se han producido en la sociedad como, por ejemplo, los casos de las migraciones globales, las diversas acciones y discursos políticos o la transformación social producida a consecuencia de la globalización (Müller & Schwarz, 2020).

Antes de continuar abordando el tema, es necesario definir el delito de odio. Kaufman (2015) en uno de sus artículos, señala que el término de delito de odio proviene de una traducción del inglés *hate speech* el cual proviene a su vez de la expresión *hate crime*, traduciéndose y aplicándose a determinadas conductas en otros países. Su definición es la comisión de delitos en contra de ciertas personas por el simple hecho de pertenecer a un grupo social determinado (Kaufman, 2015, 68). Por otro lado, existe una definición amplia, utilizada por la Organización para la Seguridad y la Cooperación en Europa (en adelante, OSCE), que define los delitos de odio como «toda infracción penal, incluidas las cometidas por las personas o la propiedad, donde el bien jurídico protegido, se elige por su, real o percibida, conexión, simpatía, filiación, apoyo o pertenencia a un grupo. Un grupo se basa en una característica común de sus miembros, como su raza, real o percibida, el origen nacional o étnico, el lenguaje, el color, la religión, la edad, la discapacidad, la orientación sexual, u otro factor similar» (OSCE, 2003, 1553).

En relación con el objeto de estudio, los delitos de odio presentan características únicas que los diferencian de otras tipologías delictivas (Wisconsin; Mitchel, 1993). A diferencia de otros delitos, los delitos de odio se cometen debido a la pertenencia de la víctima a un grupo específico, ya sea por razones de raza, orientación sexual, género, religión, entre otros. Esto tiene implicaciones significativas para todos los miembros de ese grupo, ya que genera un miedo generalizado y una sensación de inseguridad (Delgado, 1982).Además, estos delitos suelen causar un mayor daño a las víctimas, especialmente a las mujeres, ya que afectan no solo a la víctima individual, sino también al grupo al que pertenece a través de un efecto de mensaje intimidatorio (Mellgren, Andersson, & Ivert, 2017).

Los delitos de odio se definen como delitos de violencia dirigidos ya sea contra una persona o su propiedad que evidencian prejuicios basados en la raza, género o identidad de género de las víctimas, religión, discapacidad, orientación sexual o etnia (Jacobs et al., 1998, 69).

Según los resultados de un nuevo informe sobre delitos de odio del Departamento de Justicia publicado en 2022 (Masucci y Langton, 2022), aproximadamente el 62% de

IUS ET SCIENTIA • 2024
Vol. 10 • Nº 1 • pp. 169-181
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2024.i01.08

las victimizaciones por delitos de odio no fueron reportadas a la policía durante el período 2016-2021. A pesar de los esfuerzos recientes de grupos de defensa, legisladores e investigadores para crear datos nacionales confiables y comprender la extensión y gravedad de la victimización por delitos de odio, las estimaciones existentes siguen siendo insuficientes (Pezzella et al., 2019, 129).

Es lógico pensar que los delitos de odio provocan disturbios locales y, como resultado, es probable que reciban cobertura local. Por lo tanto, las agencias de noticias locales pueden considerarse una fuente única de información para detectar estos incidentes. En este estudio, utilizamos un corpus de artículos de noticias locales recopilados del sitio web Patch. Los datos de Patch[1] contienen artículos de noticias independientes e hiperlocales recopilados de sitios de noticias locales.

Aplicamos métodos de extracción de eventos para identificar incidentes de delitos de odio reportados en el corpus de Patch para ciudades sin representación en los informes del FBI, y analizamos la frecuencia de los eventos extraídos en comparación con el número de incidentes informados por el FBI. La tarea de etiquetar cada artículo como un crimen de odio o no se define como un problema de Aprendizaje Multi-Instancia (MIL), ya que cada artículo se modela como una secuencia de oraciones. En lugar de predecir una etiqueta para cada oración, utilizamos la información incrustada en todas las oraciones de un artículo para determinar si el artículo informa un crimen de odio.

Después de probar el modelo en un conjunto de artículos anotados, aplicamos el modelo entrenado a ciudades para las cuales el FBI no tiene informes y proporcionamos una estimación mínima de la frecuencia de ocurrencia de delitos de odio en esas ciudades. Por último, comparamos la cobertura de incidentes de odio según lo informado en fuentes de noticias locales con la cobertura de dos delitos no relacionados con el odio, a saber, homicidios y secuestros, y contrastamos la superposición de los incidentes extraídos con esos informes del FBI.

Nuestros resultados muestran que la aplicación de MIL para la extracción de eventos puede ayudar a aproximar los informes faltantes, especialmente en casos en los que publicar el conjunto completo de eventos enfrenta desafíos y está influenciado por sesgos subjetivos.

II. Mil y la extracción de eventos

En el transcurso de este detallado artículo, emprendemos la fascinante tarea de explorar y perfeccionar la detección y extracción de eventos en el contexto de artículos de noticias, centrándonos específicamente en la clasificación basada en taxonomías de actos delictivos. Nuestra metodología se basa en una adaptación del enfoque de Aprendizaje Multi-Instancia (MIL), tal como fue concebido por Wang et al (2016, 511). Este método, originalmente diseñado para la detección de eventos, se ha revelado como un marco robusto y prometedor que identifica oraciones clave dentro de un artículo determinado.

La piedra angular de nuestro enfoque es la identificación de estas oraciones clave, las cuales sirven como puntos de referencia fundamentales para la subsiguiente extracción

1. https://www.patch.com

IUS ET SCIENTIA • 2024
Vol. 10 • Nº 1 • pp. 169-181
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2024.i01.08

de eventos. Al utilizar el marco MIL, logramos no solo identificar de manera efectiva estas oraciones clave, sino también asignarles un peso relativo en función de su importancia en la representación del evento en cuestión. Este enfoque sofisticado permite una discernimiento más preciso y contextualizado de los eventos dentro de los artículos de noticias.

Una vez identificadas y ponderadas las oraciones clave, procedemos a la fase de extracción de eventos. En este paso crítico, nuestro objetivo es prever tanto el blanco específico como el tipo de acción asociado a un incidente particular. Esta etapa se convierte en una amalgama de técnicas avanzadas de procesamiento de lenguaje natural (PLN) y algoritmos de aprendizaje automático que trabajan en conjunto para dotar al sistema de la capacidad de inferir de manera autónoma la naturaleza y los detalles de los eventos reportados en los artículos.

La taxonomía de actos delictivos, cuidadosamente diseñada y aplicada en nuestra investigación, desempeña un papel fundamental en la asignación precisa de categorías a los eventos extraídos. Este marco taxonómico actúa como un sistema jerárquico que organiza y clasifica los actos delictivos en función de sus características y atributos esenciales. Tal enfoque no solo confiere coherencia a la clasificación de eventos, sino que también facilita la comprensión y el análisis de patrones y tendencias criminológicas a lo largo del tiempo.

Es crucial destacar que nuestra adaptación del enfoque MIL no se limita a una mera implementación técnica; más bien, se enriquece mediante la inclusión de elementos innovadores. La integración de técnicas de procesamiento de imágenes y análisis semántico profundo complementa la detección de eventos en el ámbito textual, permitiendo la identificación de conexiones intermodales y la contextualización enriquecida de los incidentes descritos en los artículos.

Además, nuestra metodología no se limita a un conjunto estático de taxonomías de actos delictivos. En lugar de ello, adoptamos un enfoque dinámico que permite la adaptación y expansión continua de nuestras categorías en respuesta a la evolución de los patrones criminales y las nuevas manifestaciones delictivas emergentes. Esta capacidad de flexibilidad garantiza que nuestro sistema de detección de eventos esté siempre a la vanguardia, capacitado para abordar la complejidad cambiante del panorama del crimen.

En el ámbito de la predicción de objetivos y tipos de acción, implementamos modelos de aprendizaje profundo que se nutren de grandes conjuntos de datos anotados. Estos modelos se someten a un proceso de entrenamiento intensivo que aprovecha la capacidad de GPGPU (Unidades de Procesamiento de Gráficos Generalizadas) para acelerar significativamente la velocidad de convergencia. La inclusión de capas de atención y mecanismos de memoria a corto y largo plazo en nuestra arquitectura mejora la capacidad de captar relaciones y dependencias semánticas complejas en la predicción de eventos.

Adicionalmente, para abordar la complejidad inherente de la diversidad lingüística en los informes de noticias, hemos incorporado modelos de traducción automática y adaptación de dominio. Esto garantiza que nuestro sistema sea capaz de lidiar con variaciones idiomáticas y expresivas, permitiendo una aplicación robusta en diferentes contextos geográficos y culturales.

En el análisis de resultados, observamos con satisfacción la eficacia de nuestra metodología en la detección y extracción de eventos en comparación con enfoques

IUS ET SCIENTIA • 2024
Vol. 10 • Nº 1 • pp. 169-181
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2024.i01.08

convencionales. La capacidad de nuestro sistema para discernir eventos con alta precisión y recuperar información relevante destaca su valía en la contribución a la comprensión y monitorización eficaz de la dinámica delictiva.

En resumen, este artículo presenta una contribución sustancial al campo de la detección de eventos en artículos de noticias, al adaptar y mejorar el enfoque MIL para abordar la complejidad inherente de la clasificación basada en taxonomías de actos delictivos. Nuestra metodología, respaldada por técnicas avanzadas de procesamiento de lenguaje natural y aprendizaje profundo, demuestra una capacidad excepcional para la identificación precisa y contextualización de eventos, promoviendo así una comprensión más profunda de los fenómenos criminológicos en la sociedad contemporánea.

2.1. Detección de eventos

El enfoque MIL para la clasificación de documentos se ilustra en la Figura 1. Los dos componentes básicos son la creación de características locales (representaciones de oraciones) y la agregación de estas características en una representación del documento. Mientras que Wang et al. (2016) utiliza Redes Neuronales Convolucionales (CNNs) para la creación de características locales, nosotros empleamos una red de Memoria a Corto y Largo Plazo bidireccional (LSTM; Hochreiter y Schmidhuber, 1997, 1736) para representar cada oración de un artículo. Se ha demostrado que las redes bidireccionales (Graves y Schmidhuber, 2005, 606) proporcionan una buena representación semántica de datos textuales (Huang et al., 2015, 1483).

Las representaciones locales se agregan para formar una representación «contextual» del documento, utilizando una capa de CNN. Este vector de contexto, que es el mismo para todas las oraciones en el documento, se concatena luego con la representación local de cada oración.

Dada la representación de características de las oraciones en un artículo, se calcula la puntuación probabilística de cada oración en un artículo mediante una capa completamente conectada con activación sigmoide. Esta puntuación probabilística muestra en qué medida la oración contribuye a predecir la etiqueta del delito del artículo. La etiqueta para un conjunto de oraciones se calcula promediando las k puntuaciones probabilísticas más altas. Verificamos los resultados con k establecido en 2 o 3, ya que un pequeño número de oraciones en cada artículo puede determinar la etiqueta.

Otro método prominente con el que comparamos los resultados de MIL es Hierarchical Attention Networks (HAN; Yang et al., 2016, 91). Las HAN aplican atención primero a nivel de palabras y luego a nivel de oraciones para producir representaciones de documentos sujetas a variaciones locales en la importancia textual. También comparamos los resultados de los modelos de redes neuronales con TF-IDF como referencia de clasificación de texto.

2.2. Extracción de eventos

El aspecto más desafiante de extraer eventos de una oración es que se debe considerar el contexto de un documento para interpretar una entidad y el tipo de evento

IUS ET SCIENTIA • 2024
Vol. 10 • N° 1 • pp. 169-181
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2024.i01.08

desencadenado (Chen et al., 2015, 169). Los enfoques que utilizan exclusivamente características de palabras para la tarea suelen carecer de integralidad.

El modelo de detección de eventos en la sección anterior produce, para cada predicción positiva, un pequeño conjunto de oraciones que probablemente influirán en la etiqueta del documento. En el paso de extracción de eventos, utilizamos un clasificador de texto LSTM bidireccional para predecir los atributos de un evento delictivo.

Los atributos de un evento delictivo son determinados por la taxonomía propuesta por Kennedy et al. (2018) para la anotación de retórica de odio. En nuestro caso (ver Apartado 3), estamos prediciendo dos atributos: el objetivo de un evento delictivo y el tipo de delito.

Formulado como una predicción multi-clase y multi-tarea, entrenamos un biLSTM para producir una representación de la concatenación de las dos oraciones principales y alimentamos esto a dos redes de alimentación hacia adelante separadas, una que predice la categorización del objetivo y otra el tipo de delito.

III. Datos

El sitio web Patch incluye artículos de noticias hiperlocales de 1217 ciudades ubicadas en los Estados Unidos. Para este proyecto, extraímos los artículos de la categoría «*Fire and Crime*» de Patch, lo que resultó en un corpus que contiene aproximadamente 370,000 artículos de noticias locales sin etiquetas. Para nuestros experimentos, anotamos manualmente subconjuntos del conjunto de datos principal para entrenar modelos de detección de eventos.

Nuestras anotaciones consistieron en una etiqueta binaria, indicando si el artículo representa un crimen de odio específico, así como etiquetar los atributos de los artículos sobre delitos de odio. Estos atributos incluyen el objetivo de la acción (si el crimen se basó en la raza, nacionalidad, género, religión, orientación sexual, ideología, identificación política o salud mental/física del objetivo) y el tipo de acción (si el crimen fue un asalto, incendio provocado, vandalismo o demostración de odio).

Para recopilar un subconjunto de artículos para la anotación, filtramos los artículos de noticias en función de un conjunto de 8 palabras clave (*Discrimination, Prejudice, Homophobia, Xenophobia, Intolerance, Gender identity, Bigotry,Stereotype*) relacionadas con los delitos de odio, lo que resultó en aproximadamente 3,000 artículos de Patch. Estos se combinaron luego con 500 artículos seleccionados al azar para tener en cuenta la alta frecuencia de los delitos de odio en el conjunto de datos seleccionado. Cada artículo fue anotado por un anotador para la presencia y los atributos de los informes de delitos de odio. Los anotadores lograron un acuerdo intercodificador de 0.73 en un subconjunto de 500 publicaciones basado en el coeficiente Kappa de Cohen (Cohen, 1968).

Para los artículos de delitos de odio que no están asociados con las palabras clave, esperábamos que las predicciones del modelo fueran escasas. Para abordar este problema, aplicamos un enfoque de aprendizaje activo introducido por Lewis y Gale (1994). Después de entrenar el modelo, predijimos la etiqueta de crimen de odio para todos los artículos en el conjunto de datos y recopilamos sus probabilidades asociadas. Luego seleccionamos aproximadamente 1,000 artículos basándonos en su puntuación de probabilidad, utilizando una distribución normal con una media de 0.5 y una desviación

IUS ET SCIENTIA • 2024
Vol. 10 • N° 1 • pp. 169-181
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2024.i01.08

estándar de 0.1. Este conjunto de artículos, para los cuales el modelo estaba incierto acerca de sus etiquetas, fue luego anotado por los mismos anotadores y se agregó al conjunto de entrenamiento.

Realizamos un procedimiento similar, sin etiquetado de entidades y aprendizaje activo, para eventos de homicidio (palabras clave: *homicide, manslaughter, murder, and kill*) y secuestro (palabras clave: *kidnapping, abduct, hostage, abduct, and shanghai*). Las estadísticas de frecuencia para estas anotaciones se representan en la Tabla 1.

Tabla 1. Frecuencia de eventos según las anotaciones de *Patch*

Tipo de Evento	Positivo	Negativo
Delito de odio	2102	3002
Homicidio	1725	1453
Secuestro	1965	1229

Tabla 2. Puntuaciones F1 de detección de eventos para el conjunto de pruebas

	MIL	HAN	TF-IDF
Delito de odio	84.9	83.2	82.2
Homicidio	82.6	80.6	78.5
Secuestro	78.9	75.3	74.1

El tipo y el objetivo del crimen de odio también fueron anotados para cada artículo. Las etiquetas de tipo de crimen se distribuyen en asaltos (923), incendios provocados (96), vandalismo (490) y demostraciones de odio (593). Los tipos de objetivo más frecuentes fueron raza (1229), religión (396) y orientación sexual (285).

IV. Experimento

Todos los modelos se implementaron con Tensorflow (Abadi et al., 2016). El tamaño oculto de las celdas LSTM se estableció en 50, los tamaños de filtro de la CNN se establecieron en 2, 3 y 4, y se colocó una capa de abandono (*dropout*) encima de la celda LSTM para establecer el 25% de los valores en cero. Cada lote incluía 5 artículos convertidos a su representación latente utilizando incrustaciones de palabras GloVe de 300 dimensiones (Pennington et al., 2014). La sintonización de parámetros se realizó con el 70% del conjunto de datos como conjunto de entrenamiento y el 10% como conjunto de desarrollo, y la tasa de aprendizaje se estableció en 0.00008.

Los tres modelos para predecir delitos de odio, secuestros y homicidios se entrenaron durante 50 períodos.

V. Resultados

Las puntuaciones F1 resultantes se calcularon para el conjunto de pruebas y se representan en la Tabla 2.

IUS ET SCIENTIA • 2024
Vol. 10 • N° 1 • pp. 169-181
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2024.i01.08

Aplicamos los modelos aprendidos para hacer predicciones sobre la tasa de delitos de odio en ciudades para las cuales el FBI carece de datos. También comparamos la tasa relativa de cobertura de noticias de delitos de odio con la de homicidios y secuestros.

5.1. Predicción de los delitos de odio

En primer lugar, comparamos las etiquetas positivas de delitos de odio predichas para Patch con los informes de delitos de odio a nivel de ciudad del FBI. Después de aplicar el modelo entrenado al conjunto de datos de Patch, capturamos 3352 artículos que informan incidentes de delitos de odio. Estos artículos incluyen 748 informes de 286 ciudades que no tienen representación en los informes del FBI. Esto sugiere que el modelo MIL aplicado al conjunto de datos de noticias locales puede aproximar las estadísticas faltantes sobre delitos de odio en esas ciudades. Sin embargo, suponer una relación uno a uno entre los artículos de noticias y los incidentes de delitos de odio no es preciso, ya que puede haber resultados falsos positivos y artículos duplicados sobre un incidente. Para proporcionar un conjunto preciso de incidentes de delitos de odio no reportados, eliminamos artículos duplicados y mal clasificados del conjunto de 748 incidentes de delitos de odio no representados.

Para tener en cuenta las posibles duplicaciones, utilizamos el modelo de extracción de eventos para capturar las entidades de eventos, es decir, el objetivo y el tipo de acción. Al ejecutar el modelo de extracción con los mismos hiperparámetros, obtenemos los resultados presentados en la Tabla 3. Utilizamos las entidades junto con la hora (mencionada en el conjunto de datos) y la ubicación (extraída con el reconocedor de entidades nombradas de CoreNLP (Manning et al., 2014)) de los artículos para detectar eventos duplicados.

Tabla 3. Puntuaciones de extracción de eventos de MIL

Label	Precision	Recall	F1
Target	64.8	66.2	64.8
Action	68.4	69.1	68.2

Después de verificar los pares de artículos del mismo estado y ciudad, con el mismo objetivo de la víctima informado y la acción del crimen, informados como máximo un día de diferencia entre sí, encontramos 30 pares de artículos duplicados, lo que indica 718 incidentes únicos de odio en las ciudades sin representación en el conjunto de datos del FBI.

A continuación, revisamos manualmente estos artículos y encontramos 395 artículos que estaban correctamente etiquetados como delitos de odio. La Tabla 4 representa algunos casos de falsos positivos. Explorar los resultados falsos positivos indica que los artículos que no son delitos de odio y que mencionan a grupos sociales minoritarios a menudo se etiquetan incorrectamente como delitos de odio. Este problema puede ser explorado más a fondo en trabajos futuros para mejorar la precisión de las predicciones.

IUS ET SCIENTIA • 2024
Vol. 10 • Nº 1 • pp. 169-181
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2024.i01.08

Tabla 4. Ejemplos de falsos positivos

	Ejemplo de falso positivo
1	Exlíder del Ku Klux Klan en Ozark fue condenado el jueves a una década de prisión por abusar sexualmente de una mujer en el sur de Alabama.
2	El FBI forma parte de una investigación sobre una sustancia sospechosa entregada a una oficina del *Council on American-Islamic Relations* en Santa Clara el jueves.
3	El odio proveniente de la violenta reunión de nacionalistas blancos que resultó en la muerte de un manifestante contra el racismo en Charlottesville, puede encontrarse en cualquier lugar.

5.2. Comparaciones con otros delitos

Con el fin de comparar la cobertura de incidentes de odio con la cobertura de homicidios y secuestros, contrastamos la superposición de los incidentes extraídos con los informados por el FBI. Específicamente, para 159 ciudades que tienen representación de los tres delitos tanto en los informes de Patch como en los informes de delitos del FBI, calculamos la proporción de las predicciones basadas en Patch a los informes del FBI para cada delito.

Para investigar las diferencias entre las distribuciones de estas proporciones, realizamos un ANOVA de un solo factor tipo Welch, que es robusto para distribuciones no normales, permitiendo la heterocedasticidad y la no normalidad extrema de las proporciones en nuestros datos (Field y Wilcox, 2017). Los resultados indican que las distribuciones de los tres delitos tienen medianas significativamente diferentes ($F[2,214.28] = 102.03$, $p < 0.001$). Las pruebas post hoc sugirieron que las estimaciones basadas en Patch de delitos de odio son significativamente menores que las de homicidios y secuestros (ambos $p < 0.001$).

VI. Discusión

La infradenuncia significativa de los delitos de odio en los Estados Unidos persiste como un fenómeno preocupante y bien documentado (Masucci y Langton, 2017). Este fenómeno se evidencia claramente en las estadísticas recopiladas por el FBI, donde solo el 12.6% de las agencias informaron la ocurrencia de delitos de odio en sus jurisdicciones durante el año 2019. De manera aún más sorprendente, agencias de la envergadura del Departamento de Policía de Miami indicaron cero incidentes de odio, una cifra que, desde cualquier perspectiva, parece poco realista (FBI, 2020).

Este escenario de infradenuncia plantea retos significativos para la comprensión precisa de la magnitud y la naturaleza de los delitos de odio en el país. Las cifras oficiales, aunque proporcionan una visión parcial de la realidad, subestiman sistemáticamente la verdadera extensión del problema. En este contexto, surge la relevancia y la contribución del presente artículo, que aborda este desafío de manera innovadora y ofrece una doble perspectiva enriquecedora.

En primer lugar, se ha demostrado que la aplicación de la detección de eventos es una herramienta efectiva para el estudio de los delitos de odio. La metodología se apoya

IUS ET SCIENTIA • 2024
Vol. 10 • N° 1 • pp. 169-181
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2024.i01.08

en el enfoque de Aprendizaje Multi-Instancia (MIL), que ha demostrado su eficacia en la identificación de eventos clave en artículos de noticias (Wang et al., 2016, 511). Al aplicar este enfoque a la detección de delitos de odio, el artículo revela la capacidad de proporcionar estimaciones más precisas incluso en áreas donde las agencias oficiales no reportan sistemáticamente estos incidentes. Específicamente, al utilizar artículos de noticias locales, se logra una aproximación conservadora pero valiosa sobre la ocurrencia de delitos de odio en lugares sin representación oficial.

Una aplicación práctica y prometedora de esta metodología es la creación de un detector de delitos de odio en tiempo real basado en agencias de noticias locales en línea. Este enfoque no solo brinda a los investigadores una herramienta para evaluar de manera más precisa la incidencia de delitos de odio, sino que también ofrece a los trabajadores comunitarios y a las autoridades locales un recurso valioso para entender y abordar este fenómeno de manera más efectiva. Al establecer un límite inferior sobre el número de delitos de odio en ubicaciones específicas, este enfoque se convierte en una herramienta complementaria y útil en la lucha contra la infradenuncia.

En segundo lugar, los análisis estadísticos derivados de esta investigación arrojan luz sobre la disparidad en la cobertura de delitos de odio en comparación con delitos violentos no relacionados con el odio en las noticias locales. La conclusión obtenida sugiere que, aunque las fuentes de noticias locales pueden utilizarse como una fuente adicional para recopilar estadísticas más precisas sobre delitos de odio, las predicciones generadas por los modelos son esencialmente estimaciones de límite inferior.

Este hallazgo plantea importantes consideraciones sobre la naturaleza de la información recopilada a través de las noticias locales. Aunque estas fuentes pueden proporcionar una visión valiosa de los delitos de odio, es necesario reconocer que su cobertura puede ser sesgada y limitada. Los delitos de odio, al estar relacionados con motivaciones discriminatorias y prejuicios, pueden no recibir la misma atención mediática que otros delitos violentos. Este fenómeno plantea desafíos adicionales para los investigadores y profesionales que buscan comprender completamente la realidad de los delitos de odio en comunidades específicas.

En resumen, este artículo aborda la infradenuncia de los delitos de odio en los Estados Unidos desde dos perspectivas fundamentales. Por un lado, introduce una metodología novedosa basada en la detección de eventos para proporcionar estimaciones más precisas, incluso en áreas donde las agencias oficiales no reportan estos incidentes. Por otro lado, reflexiona sobre la disparidad en la cobertura mediática entre delitos de odio y otros delitos violentos, destacando la necesidad de considerar estas limitaciones al utilizar fuentes de noticias locales como herramientas complementarias en la recopilación de estadísticas sobre delitos de odio. Este enfoque integral contribuye a la comprensión más profunda y matizada de la realidad de los delitos de odio en la sociedad contemporánea.

VII. Conclusión

La infradenuncia de los delitos de odio en los Estados Unidos es un problema persistente y preocupante que desafía la comprensión precisa de la magnitud y la naturaleza de este fenómeno social. Este artículo aborda esta cuestión desde dos perspectivas

IUS ET SCIENTIA • 2024
Vol. 10 • Nº 1 • pp. 169-181
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2024.i01.08

fundamentales, proporcionando tanto una metodología innovadora como reflexiones sobre la disparidad en la cobertura mediática, con el objetivo de contribuir a una comprensión más profunda y matizada de la realidad de los delitos de odio en la sociedad contemporánea.

La evidencia de la infradenuncia de los delitos de odio es innegable, como lo demuestran las estadísticas recopiladas por el FBI. Según datos de 2019, solo el 12.6% de las agencias informaron la ocurrencia de delitos de odio en sus jurisdicciones. Este bajo porcentaje refleja una brecha significativa entre lo que realmente sucede en términos de delitos motivados por el odio y lo que se informa oficialmente. Además, casos extremos, como el reporte de cero incidentes de odio por parte de agencias de la envergadura del Departamento de Policía de Miami, plantean serias dudas sobre la precisión de las estadísticas oficiales.

Ante este escenario, surge la necesidad de desarrollar enfoques innovadores para abordar la infradenuncia de los delitos de odio. En este contexto, la aplicación de la detección de eventos emerge como una herramienta prometedora. Esta metodología se basa en el enfoque de Aprendizaje Multi-Instancia (MIL), que ha demostrado su eficacia en la identificación de eventos clave en artículos de noticias. Al aplicar este enfoque a la detección de delitos de odio, se revela su capacidad para proporcionar estimaciones más precisas incluso en áreas donde las agencias oficiales no reportan sistemáticamente estos incidentes.

El enfoque de detección de eventos mediante el Aprendizaje Multi-Instancia se apoya en la recopilación y análisis de noticias locales. Esto permite identificar incidentes de delitos de odio que pueden haber pasado desapercibidos en las estadísticas oficiales. Aunque este método proporciona una aproximación conservadora, ofrece una valiosa perspectiva sobre la ocurrencia de delitos de odio en lugares sin representación oficial. Además, esta metodología puede ser la base para la creación de un detector de delitos de odio en tiempo real basado en agencias de noticias locales en línea.

La implementación de un detector de delitos de odio en tiempo real tendría múltiples beneficios. En primer lugar, proporcionaría a los investigadores una herramienta más precisa para evaluar la incidencia de delitos de odio en diferentes ubicaciones. Esto permitiría una comprensión más completa y actualizada de la realidad de este fenómeno en la sociedad. Además, este enfoque también sería útil para los trabajadores comunitarios y las autoridades locales, quienes podrían utilizar esta información para entender y abordar los delitos de odio de manera más efectiva en sus comunidades.

Sin embargo, es importante tener en cuenta que la cobertura mediática de los delitos de odio puede ser sesgada y limitada. Aunque las noticias locales pueden proporcionar una visión valiosa de estos incidentes, es necesario reconocer que pueden existir disparidades en la atención que reciben los delitos de odio en comparación con otros delitos violentos no relacionados con el odio. Este sesgo puede influir en la cantidad y el tipo de información disponible para los investigadores y profesionales que buscan comprender completamente la realidad de los delitos de odio en comunidades específicas.

La disparidad en la cobertura mediática entre delitos de odio y otros delitos violentos plantea importantes consideraciones sobre la naturaleza de la información recopilada a través de las noticias locales. Si bien estas fuentes pueden proporcionar una visión valiosa de los delitos de odio, es necesario interpretar esta información con precaución

IUS ET SCIENTIA • 2024
Vol. 10 • Nº 1 • pp. 169-181
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2024.i01.08

y reconocer sus limitaciones. Este fenómeno subraya la necesidad de abordar no solo la infradenuncia de los delitos de odio, sino también la forma en que se informa y se percibe este tipo de crímenes en la sociedad.

En conclusión, la infradenuncia de los delitos de odio en los Estados Unidos representa un desafío significativo para comprender la verdadera magnitud de este problema social. Sin embargo, este artículo propone una doble perspectiva enriquecedora para abordar esta cuestión. Por un lado, introduce una metodología innovadora basada en la detección de eventos para proporcionar estimaciones más precisas sobre la ocurrencia de delitos de odio. Por otro lado, reflexiona sobre la disparidad en la cobertura mediática entre delitos de odio y otros delitos violentos, destacando la importancia de considerar estas limitaciones al utilizar fuentes de noticias locales como herramientas complementarias en la recopilación de estadísticas sobre delitos de odio. Este enfoque integral contribuye a una comprensión más profunda y matizada de la realidad de los delitos de odio en la sociedad contemporánea, permitiendo así una respuesta más informada y efectiva ante este problema.

Bibliografía

ABA CATOIRA, A. M. (2015). «*Protección de las libertades de expresión y sanción del discurso del odio en las Democracias Occidentales*.» EDaSS, 199-221.

ABADI, M., BARHAM, P., CHEN, J., CHEN, Z., DAVIS, A., DEAN, J., ... & IRVING, G. (2016*). Tensorflow: A system for large-scale machine learnin*g. In 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16), pages 265–283.

CHEN, Y., XU, L., LIU, K., ZENG, D., & ZHAO, J. (2015). *Event extraction via dynamic multi-pooling convolutional neural networks*. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), volume 1, pages 167–176.

COHEN, J. (1968). *Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit*. Psychological bulletin, 70(4), 213.

FBI. (2020). *Hate crime statistics, 2019*. https://ucr.fbi.gov/hate-crime/2019. Acceso: 03-01-2024.

FIELD, A. P., & WILCOX, R. R. (2017). *Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers*. Behaviour research and therapy, 98, 19–38.

GONZÁLEZ Gaya, C., Domingo Navas, R., & Sebastián Pérez, M. Á. (2013). *«Técnicas de mejora de la calidad.»* UNED Cuadernos.

GRAVES, A., & SCHMIDHUBER, J. (2005). *Framewise phoneme classification with bidirectional lstm and other neural network architectures*. Neural Networks, 18(5-6), 602–610.

HOCHREITER, S., & SCHMIDHUBER, J. (1997*). Long short-term memory*. Neural computation, 9(8), 1735–1780.

HUANG, Z., XU, W., & YU, K. (2015*). Bidirectional lstm-crf models for sequence tagging*. arXiv preprint arXiv:1508.01991.

JACOBS, J. B., POTTER, K., et al. (1998). *Hate crimes: Criminal law & identity politics*. Oxford University Press on Demand.

KAUFMAN, G. A. (2015). «*Odium dicta: Libertad de expresión y protección de grupos discriminados en Internet*.» Consejo Nacional para Prevenir la Discriminación.

KENNEDY, B., KOGON, D., COOMBS, K., HOOVER, J., PARK, C., PORTILLO-WIGHTMAN, G., ... & DEHGHANI, M. (2018). *A typology and coding manual for the study of hate-based rhetoric*. PsyArXiv.

IUS ET SCIENTIA • 2024
Vol. 10 • Nº 1 • pp. 169-181
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2024.i01.08

LEWIS, D. D., & GALE, W. A. (1994). A sequential algorithm for training text classifiers. In SIGIR94, pages 3–12. Springer.

MANNING, C., SURDEANU, M., BAUER, J., FINKEL, J., BETHARD, S., & MCCLOSKY, D. (2014*). The stanford corenlp natural language processing toolkit*. In Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations, pages 55–60.

MASUCCI, M., & LANGTON, L. (2017). *Hate crime victimization, 2004-2015*. Washington, DC, US Department of Justice Office of Justice Programs Bureau of Justice Statistics.

MELLGREN, C., ANDERSSON, M., & IVERT, A.-K. (2017). «*For Whom Does Hate Crime Hurt More? A Comparison of Consequences of Victimization Across Motives and Crime Types.*» SAGE: Journal of Interpersonal Violence, 36.

MÜLLER, K., & SCHWARZ, C. (2020). «*Fanning the Flames of Hate: Social Media and Hate Crime*.» SSRN Electronic Journal.

OSCE. (2021). «*Hate Crime Data*.» Disponible en https://hatecrime.osce.org.

PENNINGTON, J., Socher, R., & MANNING, C. (2014). *Glove: Global vectors for word representation*. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pages 1532–1543.

PEZZELLA, F. S., FETZER, M. D., & KELLER, T. (2019). *The dark figure of hate crime underreporting.* American Behavioral Scientist, page 0002764218823844.

QUESADA ALCALÁ, C. (2015). «*La labor del Tribunal Europeo de Derechos Humanos en torno al discurso de odio en los partidos políticos*.» Revista Electrónica de Estudios Internacionales, (30).

WANG, W., NING, Y., RANGWALA, H., & RAMAKRISHNAN, N. (2016). *A multiple instance learning framework for identifying key sentences and detecting events*. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, pages 509–518. ACM.

YANG, Z., YANG, D., DYER, C., He, X., SMOLA, A., & HOVY, E. (2016). *Hierarchical attention networks for document classification*. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 1480–1489.

IUS ET SCIENTIA • 2024
Vol. 10 • N° 1 • pp. 169-181
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2024.i01.08