

## Uso de diccionarios en Python para conocer las tendencias en Twitter sobre el aumento de feminicidios en México

Marleni Reyes Monreal<sup>1</sup>, Beatriz Bernábe Loranca<sup>2</sup>,  
Estefania Sarmiento Barrios<sup>2</sup>, Melissa Isaaly Mendoza Bernábe<sup>3</sup>

<sup>1</sup> Benemérita Universidad Autónoma de Puebla,  
Escuela de Artes Plásticas y Audiovisuales,  
México

<sup>2</sup> Benemérita Universidad Autónoma de Puebla,  
Facultad de Ciencias de la Computación,  
México

<sup>3</sup> Universidad Iberoamericana Puebla,  
México

marleni.reyes@correo.buap.mx

**Resumen.** Debido a las noticias alarmantes en las redes y diversos medios de difusión, se han establecido distintos supuestos sobre un aumento de violencia en contra de la mujer, en particular en el asunto de feminicidios. Esta situación, ha dado lugar a la necesidad de identificar, las tendencias de opinión y forma en que se difunden y viralizan los contenidos relacionados con el tema de feminicidios en México. En este punto, la técnica de creación de diccionarios aplicados al estudio de contenidos en Twitter mostrará las relaciones entre palabras usadas para referirse al tema y el sentido general en el que se comprende el tema. Considerando que el desarrollo de una investigación centrada en un problema del tamaño de feminicidios implica un elevado costo económico, de tiempo, de recursos humanos, entre otros. Identificar y establecer justicia o solución al este problema conduce a un trabajo mayúsculo con múltiples variables, con áreas interminables de analizar y representa un elevado precio en distintas direcciones. Es en este punto donde se inserta el artículo que presenta una propuesta de creación de diccionarios respecto a comentarios en Twitter con el propósito de organizar las palabras representativas sobre feminicidios de acuerdo con las opiniones de los usuarios de Twitter, una de las redes sociales más utilizadas para compartir noticias. De esta manera, es posible proponer actividades de reflexión, difusión, prevención e incluso diagnóstico sobre la magnitud e incidencia del problema sumada a las investigaciones de otras áreas que tratan el tema y sin el elevado costo que tiene una investigación tradicional.

**Palabras clave:** Redes sociales, análisis de datos, diccionarios, feminicidios.

## Using Dictionaries in Python to Understand Twitter Trends about the Rise in Femicides in Mexico

**Abstract.** Due to alarming news stories on social media and in various media outlets, various assumptions have arisen about an increase in violence against women, particularly in the area of femicide. This situation has led to the need to identify opinion trends and the way in which content related to femicide in Mexico is disseminated and goes viral. At this point, the dictionary creation technique applied to the study of content on Twitter will reveal the relationships between words used to refer to the topic and the general understanding of the topic. Considering that conducting research focused on a problem the size of femicide entails a high cost in terms of finances, time, human resources, and other aspects. Identifying and establishing justice or a solution to this problem requires a massive undertaking with multiple variables, endless areas to analyze, and represents a high price in different directions. This is where the article presents a proposal for the creation of dictionaries of Twitter comments. This would organize representative words about femicides based on the opinions of Twitter users, one of the most widely used social networks for sharing news. This would allow for activities for reflection, dissemination, prevention, and even diagnosis of the magnitude and impact of the problem, combined with research from other areas that address the topic and without the high cost of traditional research.

**Keywords:** Social networks, data analysis, dictionaries, femicides.

### 1. Introducción

El concepto de feminicidio fue propuesto por la norteamericana Diana E.H Russell, quien lo considera como “el asesinato de mujeres a manos de hombres debido a que son mujeres” [1]. En México, los casos de asesinatos de mujeres en Ciudad Juárez Chihuahua se comenzaron a denunciar en 1993 dando lugar a que el tema se posiciona como un problema de relevancia nacional e internacional. A partir de ese momento, surgieron distintas protestas para exigir justicia y seguridad para la ciudadanía mexicana, particularmente para las víctimas; sin embargo, tuvieron que pasar casi 20 años para que el país incluyera en su código penal el delito de feminicidio [2].

El aumento considerable de la violencia en el país que se traduce en crímenes, intimidaciones, abusos, injusticias y arbitrariedades contra las mujeres, no ha sido suficiente para que las autoridades establezcan un orden y medidas adecuadas para detener la violencia de género. Una gran mayoría de los casos de feminicidio continúan sin ser formalmente registrados, investigados, juzgados o sancionados. Se asume que la ausencia de grupos especializados que trabajen de manera eficiente y transparente limita las investigaciones necesarias. Por otra parte, la falta de datos estadísticos y

sistemas de información verídicos o confiables, que contribuyan a los procesos de las investigaciones, impiden la valoración de la gravedad del problema.

De acuerdo con lo anterior, se propone un estudio adicional sobre problema que analice lo referente a su difusión y las opiniones que genera. Por tanto, el objetivo de este trabajo es realizar un estudio social y computacional de los comentarios en Twitter que permita generar datos e información para conocer el impacto en la visión del público del problema de feminicidios en México en los últimos años. Consecuentemente, al caracterizar el problema dentro de una red social, es posible estructurar estadísticamente los resultados obtenidos y proponer un programa de difusión que incida en los valores de la gente que transita en las redes sociales.

Adicionalmente, se busca identificar la presencia del tema de feminicidios respecto a opiniones, relatos, lugares y maneras que se asocian a este problema, con el fin de estructurar la información en diccionarios que contienen las palabras más representativas relacionadas con feminicidio.

El trabajo se encuentra estructurado de la siguiente manera, A) presentar algunos trabajos relacionados, B) exponer el primer acercamiento a la solución del problema usando PLN, C) crear los diccionarios, D) presentar conclusiones y trabajo futuro.

## **2. Preliminares y trabajos relacionados**

En este apartado se incluyen algunos artículos que abordan el tema del análisis sobre feminicidios usando Procesamiento de Lenguaje Natural (PLN), además de trabajos que permiten conocer los conceptos básicos como feminicidio, violencia, violencia de género y dignidad. El orden de aparición no dictamina la importancia.

### **2.1 Trabajos relacionados**

#### **Redes en Twitter y la defensa de la mujer peruana (ante el acoso y el feminicidio)**

En este trabajo del 2020 se analiza el tema de feminicidios en Perú y el uso de Twitter como forma de protección para la mujer. Para hacerlo, se recolectan Tweets, usando etiquetas/hashtags y palabras clave, se aplican técnicas de PLN para automatizar el análisis de los mensajes y se usan métricas de centralidad durante el análisis de redes para la identificación de los perfiles más relevantes. Finalmente, el objetivo principal del texto ha sido diseñar, modelar y analizar los datos de Twitter. El resultado de este trabajo consistió en identificar las personas influenciadoras y las tendencias que las personas en la red discuten sobre la defensa de los derechos de las mujeres y que deberían tener principal atención por parte de las autoridades [3].

#### **Tecnologías de la lengua para análisis de opiniones en redes sociales: Procesamiento del Lenguaje Natural**

Este trabajo trata sobre el análisis y monitorización del flujo de datos, utilizando métodos inteligentes de acceso a la información, así como el análisis de sentimientos o minería de opiniones. Mediante el desarrollo de este proyecto se trata de afrontar los desafíos como el ruido de los textos analizados con el fin de desarrollar un sistema

efectivo de minería de opiniones sobre micro textos escritos en español y gallego. El resultado del estudio es el desarrollo de relevantes en analizadores de dependencias basados en grafos, diseño e implementación de sistemas de minería de opiniones multilingües no supervisados [4].

### **Análisis de Flujos de Información de la Red Social Twitter Caso de Estudio: Impacto de las Tendencias en Redes Sociales para las Operaciones del Mercado de Acciones de la Empresa ECOPEPETROL**

El desarrollo de la investigación sobre flujos de información tiene el objetivo de analizar datos provenientes de Twitter utilizando una API de captura de flujos de información y procesando la misma con el fin de realizar categorizaciones cuantificables de la información para mejorar la toma de decisiones para la operación de valores en el mercado del grupo de acciones de la empresa Ecopetrol S.A [5]. Este estudio no aborda al feminicidio, pero aporta a la investigación de tendencias en Twitter y su uso en la toma de decisiones; sus métodos también pueden ser aplicados a la toma de decisiones con respecto al tema de feminicidios.

## **2.2 Feminicidio: Terminología**

### **a) Motivación: Sobre la violencia, dignidad y derecho de la mujer**

El significado de violencia siempre llevará a una connotación negativa y se refiere a la manifestación de actos de fuerza física desmedida, siendo este concepto el más difundido en los organismos jurídicos. Cabe aclarar que este no es el único tipo de violencia documentado.

El termino de violencia de genero ha tomado presencia en México debido aumento de casos y denuncias en los últimos años. En 1993, la declaración sobre la eliminación de la violencia contra la mujer de la Organización de las Naciones Unidad ONU defina a la violencia contra la mujer como “Todo acto de violencia basado en el género que tiene como resultado posible o real un daño físico, sexual o psicológico, incluidas las amenazas, la coerción o la privación arbitraria de la libertad, ya sea que ocurra en la vida pública o en la vida privada” [6].

En el convenio del Consejo de Europa sobre prevención y lucha contra la agresión generalizada a las mujeres, el término violencia de género se refiere a la violencia que se ejerce directamente contra la mujer porque el hecho de ser mujer [7].

El concepto de violencia de género y la lucha contra su erradicación ha cobrado importancia recientemente, por múltiples factores como el aumento de su difusión en redes. Aunque se estima que es alrededor de los 90s cuando las iniciativas para la eliminación de la violencia contra la mujer se consolidan.

Al estudiar los documentos sobre el tema destaca la palabra dignidad, la cual se encuentra relacionada directamente al tema, valor necesario en los derechos a la mujer [9]. Por el lado de la ética, se dice que la dignidad se sustenta en la “moral” que da forma a la conciencia de la responsabilidad que se orienta a una regencia sobre las acciones propias, por tanto, es el resultado de la manera de comportarse “íntegramente”, por encima de todo precio [10]. En este aspecto, es posible afirmar que la dignidad, no es aislada del respeto, consideración y valoración a la mujer, por lo cual, decir que la

mujer se dignifica, no es un asunto de mera capacidad reproductiva y por tanto es necesario aplicar esfuerzos para contrarrestar actitudes discriminatorias [11].

#### **b) Sobre herramientas y técnicas de Procesamiento en Lenguaje Natural.**

El Lenguaje Natural (LN) “es el medio que se utiliza de manera cotidiana para establecer comunicación entre personas” [12]. En otras fuentes, se afirma que el LN ha evolucionado para establecer la comunicación entre seres humanos [13]. Dos definiciones sencillas que sugieren que, a pesar de que el medio de comunicación puede resultar simple, este se ha vuelto un proceso que involucra distintas herramientas que van desde las intuitivas hasta las sistemáticas. El LN ha ido creciendo debido a las exigencias de las plataformas para interpretar y analizar comunicación en distintos medios.

#### **Procesamiento del lenguaje natural**

El Procesamiento de Lenguaje Natural (PLN) es la rama de las ciencias computacionales encargada del diseño e implementación de los elementos de software y hardware necesarios para el tratamiento computacional del LN. Su objetivo es la comprensión del LN por parte de la computadora [14]. Este campo ha evolucionado los últimos años debido al gran volumen de datos disponible y a la capacidad de utilizar algoritmos nuevos para su procesamiento. Por otra parte, el PLN se incluye dentro de ciencias de la computación e Inteligencia Artificial (IA), del mismo modo se incorpora en Ingeniería del conocimiento, que se entiende como el “conjunto de conocimientos y técnicas que permiten aplicar el saber científico a la utilización del conocimiento (entendimiento, inteligencia o razón natural)” [15].

#### **Twitter**

Las redes sociales permiten la relación entre usuarios, quienes comparten, difunden y transforman información, cabe resaltar que son servicios basados en web [16]. Debido a esta nueva tecnología se ha generado una gran cantidad de información que es compartida por miles de usuarios, generando a su vez más información. La gran cantidad de información y usuarios en esta red impacta en la construcción de códigos culturales, lo que hace necesario entender que el procesamiento de este volumen de información en ayuda en el análisis de diversas situaciones en distintos contextos que son compartidos por la red. La red social de interés en este documento es Twitter, entendido como un servicio que permite compartir pensamientos, enlaces, información, etc. en formato de microblog, usando tweets que son pequeñas publicaciones basadas en texto de no más de 280 caracteres. Sobre la interacción, Twitter permite la interacción con cualquier usuario web, que pertenezca a la red social, así como comunicarse de forma privada o pública [3]. Twitter es una red en tiempo real que posibilita la interacción con información que está sucediendo en el momento, instantáneamente.

Twitter posee una serie de características que permiten su aplicación dentro del proceso de análisis y extracción de información con el fin de obtener conocimiento. Los mensajes cortos con base en texto de Twitter la hacen excelente red social para ser estudiada PLN.

## Python

Para este estudio se empleó el lenguaje de programación interpretado y multipropósito Python. Este lenguaje de *scripting*, dinámico, funcional y de alta precisión numérica. Debido a su popularidad, se puede aplicar y encontrar en diversas áreas que requieren desarrollo, “se trata de un lenguaje potente, flexible, con una sintaxis muy limpia que favorece un código legible” [17].

## Feminicidio en PLN

Los feminicidios son asesinatos de mujeres por razones de género, debido a las consecuencias de estructuras de poder en un sistema patriarcal [3]. La palabra feminicidio se suele adaptar a contextos culturales y sociales, por lo que si bien, existe una idea general del concepto, también existen muchas formas de expresarla, entenderla y legislarla. Este problema ha sido analizado con diversas herramientas computacionales incluyendo PLN.

### 3. Primera etapa: Descubrimiento de información y primer acercamiento a la solución del problema

Twitter permite exponer noticias y opiniones de un tema en tiempo real que pueden ser medidas en un tiempo determinado. Los datos de Twitter son únicos y se pueden extraer a partir de la información compartida en otras plataformas sociales

#### 3.1 Arquitectura del problema de feminicidios en PLN

La arquitectura del problema que tratamos en este artículo induce a señalar una serie de pasos a seguir para tener una implementación deseable. En términos amplios, se exhiben las indicaciones en secuencia iniciando con el descubrimiento de la información y terminando con la conclusión de los resultados (ver Fig. 1).

#### 3.2 Selección de palabras clave y hashtags

El paso previo para la extracción de datos es plantear las palabras clave que describen al problema para facilitar las descargas de los Tweets y mantener una relación constante del tema que se aborda. La manera más útil consiste en buscar las cuentas de aquellas personas influenciadoras relacionadas con el tema de feminicidios. Una vez encontradas las cuentas asociadas, se identifican las palabras más representativas del problema y se realizan búsquedas con esas palabras a través de hashtags. Con este procedimiento, las palabras con más énfasis en cuanto a la frecuencia son las siguientes: #NiU-naMas, #NiUnaMenos, #Feminicidio, #VivasNosQueremos, Acoso sexual, Movimiento feminista, Asesinato machista, Feminicida.

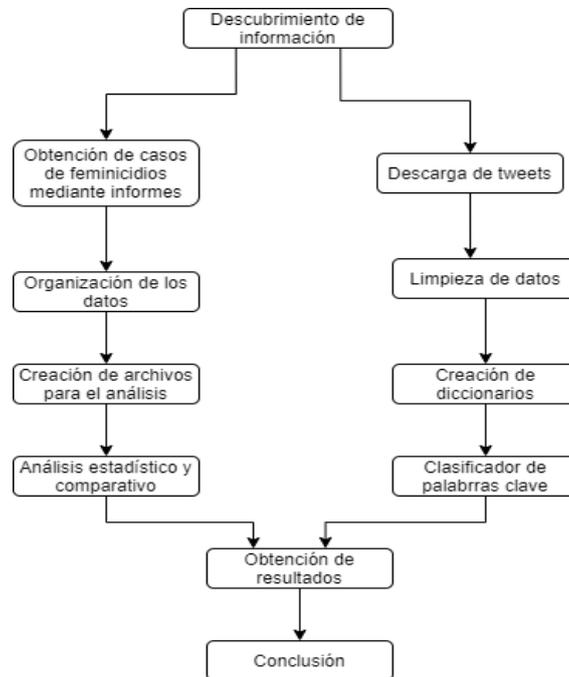


Fig. 1. Arquitectura del problema de femicidios.

### 3.3 Extracción de datos con Vicinitas

El procedimiento general consistió en descargar Tweets iniciando con 10 días consecutivos para continuar con descargas semanales. Las palabras clave se identificaron como #NiUnaMas, # NiUnaMenos, #Femicidio, #VivasNosQueremos, Acoso sexual, Movimiento feminista, Asesinato machista y Femicida. Estas palabras permitieron la extracción de los Tweets que son de interés. La limpieza de datos es un proceso necesario en la información extraída y que debe atenderse de manera implícita en las descargas debido a que muchas palabras son conectores o abreviaturas y además diversos Tweets no son relevantes para el problema por su bajo contenido significativo en femicidios.

En este escenario, al principio del descubrimiento de información en Twitter, se centró la atención en las descargas iniciales de los Tweets por 10 días consecutivos para analizar la información y continuar con descargas semanales. Para este propósito se utilizó una aplicación en la web llamada Vicinitas dado que proporciona datos básicos de los tweets del problema a explorar [18]. En un segundo paso, dos usuarios independientes analizaron los tweets principales (se excluyeron los re-tweets y las respuestas) para un análisis de contenido cualitativo y el desacuerdo se resolvió mediante la consulta de un tercer revisor independiente para señalar los tweets relevantes.

En este trabajo, la primera parte de descargas se realizó con Vicinitas para experimentar su utilidad, sin embargo, genera repetición de tweets por semana, esta situación dio lugar a controlar las descargas por día utilizando la API de Python.

Para utilizar el API de Twitter se necesita la librería tweepy de Python. También se requiere una cuenta en Twitter como desarrollador para dar de alta una aplicación y conseguir las credenciales de acceso/Tokens a las API de Twitter. Este acceso generalmente es desde un procedimiento en Python que permite las descargas de datos que se asocian a la búsqueda de palabras claves del tópico a tratar. El procedimiento posterior consiste en usar filtros de limpieza de datos y crear los archivos Excel de tipo .csv con los datos resultantes. El pseudocódigo del programa utilizado se presenta a continuación.

```
1. Inicio
2. Asignar credenciales
3. Escribir palabras a buscar
4. Leer palabra a buscar
5. Escribir número de tweets a capturar
6. Leer número de tweets
7. Crear archivo de salida
8. Obtener fecha y hora
9. Asignar nombre de archivo con palabra, fecha y hora
10. Inicializar tweets descargados como 0
11. Mientras tweets descargados < tweets a capturar hacer
    Obtener tweet
    Si tweet contiene 'áéíóú' entonces
        Cambiar valor por 'aeiou'
        Continuar
    Sino
        Continuar
    Fin si
    Si tweet contiene url entonces
        Sustraer url
        Continuar
    Sino
        Continuar
    Fin si
    Si tweet tiene caracteres especiales entonces
        Sustraer caracteres especiales
        Continuar
    Sino
        Continuar
    Fin si
    Escribir tweet en el archivo creado
    Aumentar el valor de tweet descargado
    Repetir
12. Fin
```

### 3.4 Creación de diccionarios y justificación de la exclusión de análisis de sentimientos

Para la creación de diccionarios se desarrolló una aplicación en Python que contabiliza las palabras en los archivos csv generados por la aplicación Vicinitas de Python. El programa realiza el conteo de la frecuencia de cada palabra que contengan

**Tabla 1.** Cifras de casos de presunto feminicidio dadas por el INEGI [18]

ESTADO	EN	FE	MA	AB	MA	JU	JU	AG	SE	OC	NO	DI	TOTA
	E	B	R	R	Y	N	L	O	P	T	V	C	L
AGUASCALIENTES	6	0	1	0	0	0	0	4	1	0	5	0	17
BAJA CALIFORNIA	12	0	0	3	14	19	23	7	3	0	2	10	93
BAJA CALIFORNIA SUR	1	0	1	4	2	0	0	1	0	1	0	0	10
CAMPECHE	0	0	0	0	1	0	1	1	0	0	8	3	14
COAHUILA	1	4	3	3	5	10	2	1	1	1	6	2	39
COLIMA	0	0	0	0	4	4	4	7	0	0	0	2	21
CHIAPAS	0	0	0	1	5	2	3	1	1	1	1	4	19
CHIHUAHUA	14	5	9	9	18	24	28	16	9	9	8	20	169
DISTRITO FEDERAL	4	0	1	6	18	10	14	4	3	1	1	13	75
DURANGO	2	0	3	0	3	3	2	1	0	1	1	0	16
GUANAJUATO	4	6	1	8	24	17	31	28	29	30	32	28	238
GUERRERO	6	5	0	0	9	18	20	3	0	1	9	13	84
HIDALGO	4	0	0	7	2	5	6	1	0	1	0	3	29
JALISCO	6	0	1	5	10	19	10	4	1	2	3	11	72
MÉXICO	13	13	14	22	18	13	21	10	12	9	18	19	182
MICHOACÁN	2	0	0	2	10	13	15	6	2	0	2	7	59
MORELOS	2	0	2	2	7	8	7	1	2	1	2	3	37
NAYARIT	0	0	0	0	3	2	1	1	0	0	1	0	8
NUEVO LEÓN	3	3	2	7	5	4	14	11	4	6	1	8	68
OAXACA	2	0	1	5	12	17	3	1	0	0	2	9	52
PUEBLA	3	3	1	2	9	12	12	2	3	8	5	7	67
QUERÉTARO	0	0	0	0	3	4	1	2	0	0	2	0	12
QUINTANA ROO	2	0	1	2	3	3	5	2	2	0	4	5	29
SAN LUÍS POTOSÍ	1	0	2	9	3	6	1	3	0	1	2	11	39
SINALOA	6	2	0	1	2	5	10	4	6	1	2	3	42
SONORA	5	3	5	3	6	10	3	5	7	2	11	6	66
TABASCO	0	1	2	4	5	9	4	1	1	0	2	3	32
TAMAULIPAS	2	0	0	2	6	1	2	2	1	2	1	5	24
TLAXCALA	1	0	0	0	2	7	1	1	0	0	5	1	18
VERACRUZ	5	3	0	12	10	22	17	5	0	0	1	9	84
YUCATÁN	0	0	0	1	2	0	0	0	0	0	0	1	4
ZACATECAS	4	1	1	7	6	3	9	2	6	3	2	3	47
<b>TOTALES</b>	<b>111</b>	<b>49</b>	<b>51</b>	<b>127</b>	<b>227</b>	<b>27</b>	<b>27</b>	<b>138</b>	<b>94</b>	<b>81</b>	<b>139</b>	<b>20</b>	<b>1766</b>
						0	0					9	

más de 5 letras, las palabras con menos de 5 caracteres se excluyen porque no son significativas para el estudio debido a que son preposiciones o bien palabras fuera del contexto. No se incluyen siglas ni acrónimos en el estudio. En general las palabras relacionadas con feminicidios son mayor o igual a 5 letras. Posteriormente se genera un archivo Excel para el tratamiento de los registros de las palabras y la frecuencia asociada.

Los datos se analizaron por semana para efectos de comparación sobre la intensidad de las palabras, frecuencia, ausencia o aparición de una nueva palabra, la primera semana del 15 al 21 de marzo y la segunda semana del 22 al 27 de marzo del año en curso.

El Análisis de Sentimientos AS también conocido como minería de opinión, un término muy discutido, pero no del todo comprendido, es muy útil en la monitorización de las redes sociales porque permite proporcionar una idea de la opinión pública sobre temas de interés. Para el trabajo que se aborda en este documento, se desarrolló AS, sin embargo, no fue posible dividir el conjunto de palabras en sentimientos positivos o



## **4. Estudio de estadísticas de casos de feminicidios y creación de diccionarios**

Con la finalidad de entender el problema base y la posterior correlación con lo expuesto en la red social se estudiaron las estadísticas de los casos de feminicidios en México a través de los datos del Instituto Nacional de Estadística, Geografía e Informática INEGI. El mayor número de muertes contabilizadas extraoficialmente son etiquetadas como “homicidios” con la etiqueta del género de la persona que falleció.

El repositorio del INEGI concentra de manera oficial entre 1990 a 2018, cifras de asesinatos en una base de datos, pero no especifica como feminicidios las muertes a mujeres cuando el crimen es por cuestión de género (dada la dificultad de señalar un feminicidio). En general, las bases de datos oficiales no presentan información concluyente a las muertes registradas de mujeres, ni como asesinatos o “feminicidios”, por tanto, no es posible categorizar “la clase de muertes de feminicidios”, por ello la importancia de acudir a otras herramientas, tanto computacionales como manuales, para clasificar esos casos. En este punto, es posible analizar los homicidios ocurridos a mujeres en distintos rangos y características con diccionarios a través de descargas en Twitter.

Para este trabajo y para efectos de comparación con los diccionarios PLN, fueron considerados los “homicidios” de mujeres del INEGI por mes en cada uno de los estados de la República. La Tabla 1 muestra el concentrado de datos según cifras del INEGI.

### **Herramientas para descargas en Twitter y descripción de diccionarios: Tweepy y API Search**

#### **4.1 Creación de Diccionarios**

Un diccionario es una colección no ordenada de objetos cuya estructura es empleada para almacenar información de toda clase. Para este artículo, los diccionarios tienen una importancia sustancial para el análisis estadístico descriptivo y también para identificar las palabras de mayor presencia en el tema de feminicidios.

El procedimiento aquí consiste en abrir diferentes archivos para enumerar la frecuencia de cada una de las palabras almacenadas en los archivos y consecuentemente crear un archivo .csv que contenga la palabra vinculada al número de frecuencia.

Por otra parte, se recurrió a la generación de una Wordcloud, para la creación de una nube de palabras desarrollada en Python. Este programa hace uso de distintas librerías como Wordcloud, de útil conveniencia para la creación de ilustraciones, donde destaca el análisis textual de datos. El resultado de este script fue una imagen con las palabras que más representan el tema, de esta forma se puede tener una idea aproximada del contexto que abarca el tema “feminicidios”. Las palabras se muestran la Fig. 2.

Cada una de las palabras resaltadas en la Figura 2 han sido consideradas por las bondades que suministran al programa de diccionarios. En la siguiente Tabla 2 se muestran las 10 palabras más notificadas en todos los datos recolectados, así como su respectiva frecuencia.



Fig. 3. Gráfico asociado a la tabla 2.



Fig. 4. Mapa de femicidios de María Salguero.

La periodicidad de los hashtags responde al conteo de 30 días para crear el diccionario final como se observa en las tablas 2 y 3. En las tablas las palabras se mantienen sin acentos como se encontraron en Twitter. En dichas tablas, todas las columnas son las palabras mas representativas que conforman el diccionario de femicidios. En la siguiente figura se señalan las palabras más insistentes:

Es interesante resaltar que el tiempo real en las redes posiciona algunos casos en particular en el interés del público como como Nicole, Ivana y Sofía, tres nombres de mujeres víctimas de femicidio del periodo estudiado. Estos nombres desaparecen y aparecerán otros en las redes. Adicionalmente, las palabras más usadas con relación al femicidio son: víctimas, mujer y demasiadas, y los hashtags NiUnaMas, NiUnaMenos.

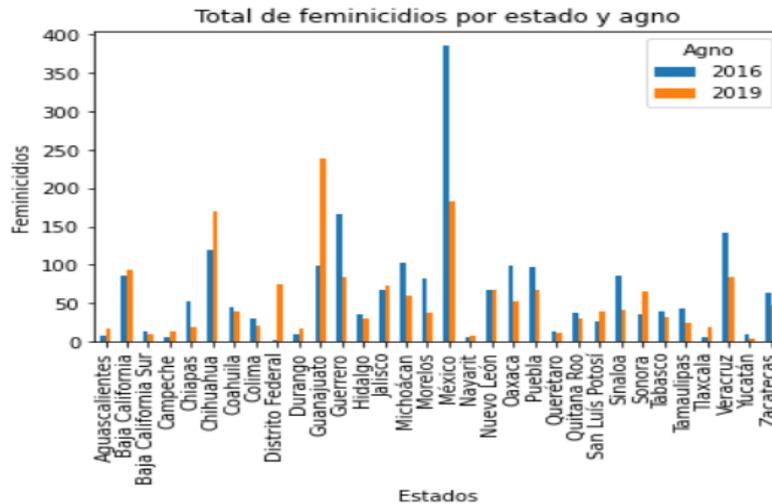


Fig. 5. Casos de feminicidios por estado en los años 2016 y 2019.

### Obtención de casos de feminicidios

Una vez creados los diccionarios y sumados a la información obtenida del INEGI, se compararon los resultados con el mapa publicado por María Salguero, en el cual se documenta mediante informes aprobados los casos de feminicidios por año en todo el país [19]. La figura 4 revela gráficamente diversas categorías, pero no existe una manera exacta de saber el número de casos de feminicidios por estado. En este punto, se contabilizaron los casos registrados por estado entre 2016 y 2019.

### Resultados: Casos de feminicidios

Se utilizó la librería “pandas” de Python para ordenar y ajustar la información del INEGI que se presentó en la tabla 1. Como consecuencia se generó la información contenida en la tabla 4 donde se revelan los resultados agrupados por año y por estado, lo cual pone en evidencia de forma tabular el aumento de feminicidios.

Se observa una inconsistencia en los feminicidios registrados con baja presencia, y la respuesta no es tan simple. Los feminicidios realmente son cifras bajas y cuando sucede uno de ellos no se califica inmediatamente como feminicidio. Esta situación se rodea de la investigación profunda para señalar como feminicidio un descenso de una mujer, generalmente se registran con otro nombre.

La Tabla 4 refleja el número de feminicidios registrados en cada estado en 2016 y en 2019. El estado con mayor número de feminicidios en 2016 fue el estado de México y los menores fueron Campeche y Tlaxcala con 5. Del mismo modo para 2019 (en la tabla aparece el distrito federal por separado, pero se sumarán sus datos al estado de México dando un total de 388 casos). Del mismo modo para 2019, el estado con mayor número de feminicidios fue Guanajuato con 238 y el menor fue Nayarit.

**Tabla 3.** Medidas de tendencia central.

	FEMINICIDIOS EN 2016	FEMINICIDIOS EN 2019
CONTEO	32.0000	32.0000
PROMEDIO	65.0625	55.1875
DESV. STD.	72.7248	53.2640
MINIMO	2.0000	4.0000
25%	14.0000	18.7500
50%	44.0000	39.0000
75%	88.7500	69.0000
MÁXIMO	386.0000	238.0000

**Tabla 4.** Femicidios por estado en los años 2016 y 2019.

ESTADO	2016	2019	ESTADO	2016	2019
AGUASCALIENTES	8	17	MICHOACÁN	102	59
BAJA CALIFORNIA	86	93	MORELOS	82	37
BAJA CALIFORNIA SUR	14	10	NAYARIT	6	8
CAMPECHE	5	14	NUEVO LEÓN	68	68
COAHUILA	45	39	OAXACA	99	52
COLIMA	30	21	PUEBLA	97	67
CHIAPAS	52	19	QUERÉTARO	14	12
CHIHUAHUA	119	169	QUINTANA ROO	37	29
DISTRITO FEDERAL	2	75	SAN LUÍS POTOSÍ	27	39
DURANGO	10	16	SINALOA	85	42
GUANAJUATO	99	238	SONORA	36	66
GUERRERO	166	84	TABASCO	39	32
HIDALGO	36	29	TAMAULIPAS	43	24
JALISCO	68	72	TLAXCALA	5	18
MÉXICO	386	182	VERACRUZ	142	84

En la Figura 5, se puede notar que, en 2016 los estados de México, Veracruz y Guerrero registraron más feminicidios.

Sin embargo, para el año 2019 el gráfico muestra que los estados en donde se registraron más casos de feminicidio son Guanajuato, el estado de México y Chihuahua, y también se distingue una disminución de los casos registrados.

Con esta información podemos concluir que en el año 2019 menos feminicidios se denunciaron y aprobaron con ese nombre. Estos resultados hacen suponer que descenso obedece a que se popularizó la denuncia formal y acusación popular contra la violencia de género gracias a la capacidad expresiva de las redes sociales, lo que provocó una alerta en que las autoridades para atender casos de violencia y evitar un feminicidio.

Por otra parte, se considera que tanto en campañas de sensibilización como en la conciencia en las medidas de castigo fueron diversificadas y extrapoladas, lo cual fomentó la prevención de feminicidios.

La información de la Tabla 4 afirma que en el año 2016 ocurrieron por estado un promedio de 65 feminicidios, mientras que en el año 2019 se redujo a 55. También se observa que en 2016 el mínimo número de casos registrados es de 2 y el máximo de 386, mientras que en el año 2019 el mínimo de casos registrados fue de 4 y el máximo de 238.

## **5. Conclusiones**

Podemos concluir que el estudio de los datos orientados a un aspecto social es muy importante porque con un orden y análisis consecuente, es posible encontrar patrones, realizar investigaciones, dar respuesta a preguntas, entre otros. Lo importante es saber obtener esta información para después transformarla en conocimiento.

Para el problema de feminicidios en particular, realizar un estudio de lo sucedido en México se considera de sustancial importancia porque se consiguió reconocer que no se le da el debido seguimiento a este delito, se habla de casos y noticias particulares y se difunden las noticias y comentarios sobre el tema masivamente, aunque no de fuentes confiables.

Con ayuda del conteo de los casos de feminicidios obtenidos del INEGI, se identificó que la violencia de género era muy común, entonces, el número de mujeres muertas no era considerado tan alarmante. Sin pérdida de generalidad, con el tiempo, estas muertes tomaron más importancia porque el número fue creciendo considerablemente. A partir de ello fue de interés el uso de las redes sociales que proporcionan generosa información pública sobre opiniones. Décadas atrás, cuando estos medios no eran tan utilizados, se había establecido la conjetura de que los feminicidios aumentaron debido a que la información no se extrapolaba y por tanto no se conocían ni la cantidad ni la magnitud del problema. El PLN ha contribuido a encontrar información y tendencias en la gran cantidad de mensajes de twitter y otras redes, para entender problemas sociales como el que se ha abordado en este trabajo. Se desea que los resultados sean difundidos a las autoridades competentes para la toma de decisiones.

Cabe subrayar que la información extraída por las redes sociales, permiten establecer el problema central con la alternativa de subrayar las palabras que lo definen, y consecuentemente sea posible crear diccionarios con las palabras de mayor ocurrencia. Estas palabras se identifican con el descubrimiento de información utilizando descargas de la red social como es el caso de este trabajo, donde diariamente se re-copilaron Tweets asociados al tópico de feminicidios y así, se crearon los respectivos diccionarios con palabras útiles al tema. Con este procedimiento ha sido posible contar con información sobre denuncias y delitos diversos sobre feminicidios, de tal modo que dependencias legales tendrán a su disposición información calibrada y estructurada que obedece a un procesamiento computacional.

## **6. Trabajo futuro**

A partir de las conclusiones ya mencionadas y observando los resultados alcanzados, es posible identificar que este tipo de temas requiere de un seguimiento e investigación

extensa y constante para obtener resultados mucho más precisos, por ello se han considerado algunos cambios en la creación de diccionarios para poder desarrollar un análisis de sentimientos consistente.

Principalmente es importante la recolección de datos con la API REST empresarial de Twitter para facilitar la extracción de la información en distintos periodos de tiempo y sin restricciones.

Para los datos estadísticos de los casos registrados, es necesario considerar más variables que ayuden a describir los factores más comunes presentes en este tipo de delito para que ayuden a una mejor interpretación. Por otra parte, se plantea considerarse la combinación de los programas de creación de diccionarios y de Wordcloud, para poder realizar el proceso de manera simultánea.

Siendo optimistas, el uso de estas redes teóricamente permite empatía y unión, lo que presume una mayor difusión del problema, y al mismo tiempo establece una reflexión en las personas.

## Referencias

1. Russell, D.; Harmes, R.; Lagarde, M.: Definición de feminicidio y conceptos relacionados. *Feminicidio: una perspectiva global*. Centro de Investigaciones Interdisciplinarias en Ciencias y Humanidades, Universidad Nacional Autónoma de México (CEIICH - UNAM), pp. 73–96 (2006)
2. Denis, D., Rodríguez, A.: Especial: Feminicidios en México, una epidemia en aumento. El país. (2020) <https://elpais.com/especiales/2017/feminicidios-en-mexico/#>.
3. Reyes Silva, P.G.: Redes en Twitter y la defensa de la mujer peruana (ante el acoso y el feminicidio). Universitat Oberta de Catalunya, repositorio institucional (2020) <http://hdl.handle.net/10609/104066>.
4. Vilares, M., Trigo, E.S., Gómez Rodríguez, C., Alonso, M. A.: Campus de Elvina, A.: Tecnologías de la lengua para análisis de opiniones en redes sociales. *Procesamiento del Lenguaje Natural* (2020) <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/5502>.
5. Sarmiento Pacanchique, E. L., Silva Téllez, D. C.: Análisis de Flujos de Información de la Red Social Twitter Caso de Estudio: Impacto de las Tendencias en Redes Sociales para las Operaciones del Mercado de Acciones de la Empresa ECOPETROL. Repositorio Institucional Universidad Distrital – RIUD. <http://repository.udistrital.edu.co/bitstream/11349/5911/1/TesisAnalisisDeFlujosDeInformacionDeLaRedSocialTwitter.pdf>.
6. ONU: Declaración sobre la eliminación de la violencia contra la mujer. Naciones Unidas Derechos Humanos, oficina del alto comisionado (1993) <https://www.ohchr.org/sp/professionalinterest/pages/violenceagainstwomen.aspx>.
7. COE: Convenio del Consejo de Europa sobre prevención y lucha contra la violencia contra las mujeres y la violencia doméstica, Serie de Tratados del Consejo de Europa – n° 210. Concil of Europe (2011) <https://rm.coe.int/1680462543>.
8. Abreu Maqueda, M.L.: La violencia de género: Entre el concepto jurídico y la realidad social. *Revista electrónica de ciencia penal y criminología* (2006) <http://criminet.ugr.es/recpc/08/recpc08-02.pdf>.
9. Martínez Bullé-Goyri, V.M.: Reflexiones sobre la dignidad en la actualidad. *Boletín Mexicano de Derecho Comparado*. (2013) <https://www.elsevier.es/es-revista-boletin>

- mexicano-derecho-comparado-77-articulo-reflexiones-sobre-dignidad-humana-actualidad-S0041863313711219.
10. Sepulveda López, M.: La dignidad humana como un valor ético jurídico implicado en la bioética y el bioderecho. *Misión Jurídica: Revista de derecho y ciencias sociales* (2009) <https://www.revistamisionjuridica.com/la-dignidad-humana-como-valor-etico-juridico-implicado-en-la-bioetica-y-el-bioderecho/>.
  11. Gil Millán, M.: La violencia sexual como un atentado contra la dignidad de la mujer. *Revista de Derecho UNED* (2015) [http://e-spacio.uned.es/fez/eserv/tesisuned:Derecho-Mgil/GIL\\_MILLAN\\_Maximiliana\\_Tesis.pdf](http://e-spacio.uned.es/fez/eserv/tesisuned:Derecho-Mgil/GIL_MILLAN_Maximiliana_Tesis.pdf).
  12. López Takeyas, B.: *Lenguaje Natural*. (2017) [http://www.itnuevolaredo.edu.mx/takeyas/Apuntes/Inteligencia%20Artificial/Apuntes/tareas\\_alumnos/PLN/PLN\(2005-II\)](http://www.itnuevolaredo.edu.mx/takeyas/Apuntes/Inteligencia%20Artificial/Apuntes/tareas_alumnos/PLN/PLN(2005-II)).
  13. Brookshear, J.G.: *Teoría de la Computación: lenguajes formales, autómatas y complejidad*. México: Pearson Educación (1993)
  14. Martín Vide, C.; Mitrana, V.; Paun, G.: *Formal Languages and Applications*. Springer (2004)
  15. Palma, J.T.; Paniagua-Aris, E.: *Ingeniería del Conocimiento. De la Extracción al Modelado de Conocimiento. Inteligencia Artificial* (2000)
  16. Orihuela, J.: *Internet: la hora de las redes sociales*. Nueva Revista de política, cultura y arte. (2008). [http://mccd.udc.es/orihuela/documentos/nueva\\_revista\\_08.pdf](http://mccd.udc.es/orihuela/documentos/nueva_revista_08.pdf).
  17. González Duque, R.: *Introducción - ¿Qué es Python? Python para todos*, pp.7 (2010)
  18. *Defunciones por homicidios*. INEGI (2020) <https://www.inegi.org.mx/sistemas/olap/proyectos/bd/continuas/mortalidad/defuncioneshom.asp?s=est>.
  19. Salguero, M.: *Los Femicidios en México* (2020) <https://femicidiosmx.crowdmap.com/FemicidiosMx>.